Indices

AIIE TRANSACTIONS

Volume 9 — 1977

Subjects (coded) and titles

0100 ***Costs***

0102 Cost Controls

"A Multi-Resource Project Scheduling Algorithm," E. M. Dar-El, Yigal Tur; March 1977, page 44.

"On the Use of A Vehicle Routing Algorithm for the Parallel Processor Problem with Sequence Dependent Changeover Costs," R. G. Parker, R. H. Deane, R. A. Holmes; June 1977, page 155.

0103 Cost Estimating

"The Transfer of Domestic-Market Production to a Foreign Site," James V. Jucker; December 1977, page 321.

0105 Productivity

"An Example of Simulation to Improve Plant Productivity," Walton Hancock, Richard Dissen, Alan Merten; March 1977, page 2.

"Dual-Phase Simulation–A Mixed Time Frame Systems Analysis Tool," David F. Gibson, Edward L. Mooney; December 1977, page 338.

0200 ***Employee Relations***

0300 ***Engineering Economics***

"Truncated Linear Gaussian Loss Integrals," James R. Buck; June 1977, page 170.

0301 Cash Flow Analysis

"A Closed-Form Methodology for Computing Present Worth Statistics of Risky Discrete Cash-Flows," James R. Buck, Jose M. A. Tanchoco; September 1977, page 278.

0304 General

"The Transfer of Domestic-Market Production to a Foreign Site," James V. Jucker; December 1977, page 321.

0400 ***Facilities Planning, Design***

0401 Environment Protection

"A Practical Method for Industrial Heat Stress Allowance Determination," Walton M. Hancock, Don B. Chaffin; June 1977, page 144.

0402 Facility Location

"A Survey of Recent Results for a Class of Facilities Location Problems," L. F. McGinnis; March 1977, page. 11.

"Extensions of the Multi-Period Facility Phase-Out Model: New Procedures and Application to a Phase-In/Phase-Out Problem," Gary M. Roodman, Leroy B. Schwarz; March 1977, page 103.

"The Rectilinear Distance Location-Allocation Problem," Anis D. Sherali, C. M. Shetty; June 1977, page 136.

"An Application of Facilities Location Theory to the Design of Forest Harvesting Areas," Dennis P. Dykstra, James L. Riggs; September 1977, page 270.

0405 Plant Layout Analysis

"Efficient Points in Location Problems," Richard E. Wendell, Arthur P. Hurter, Jr., Timothy J. Lowe; September 1977, page 238.

0500 ***Functional Operations***

0504 Maintenance, Plant Engineering

"Maintenance Scheduling for Multicomponent Equipment," Roger C. Vergin, Michael Scriabin; September 1977, page 297.

"Noise Analysis by iso-dbA Mapping," Chester J. Kishel, Edward J. R. Hudec; December 1977, page 377.

0600 ***Human Engineering***

0606 Personnel Scheduling

"A Scheduling Algorithm for the Unbalanced Production Line," Russell E. Jacobson, Randall F. Sadowski; March 1977, page 36.

0607 Personnel Selection

"Sensitivity of Optimum Workforce Size to Employee Absenteeism in Stochastic Workload Situations," Phiroze J. Nagarvala, Shlomo Globerson; December 1977, page 414.

0608 Safety and Health

"A Practical Method for Industrial Heat Stress Allowance Determination," Walton M. Hancock, Don B. Chaffin; June 1977, page 144.

"Noise Analysis by iso-dbA Mapping," Chester J. Kishel, Edward J. R. Hudec; December 1977, page 377.

0700 ***Incentive Plans***

0800 ***Industrial Engineering***

"Education and Professional Practice Criticism and Feedback," William T. Morris; September 1977, page 260.

0804 Education

"Education and Professional Practice: Criticism and Feedback," William T. Morris: September 1977, page 260.

0900 ***Industries***

0903 Health Care and Other Services

"Scheduling Multispecialty Clinics Using Integer Programming," Thom J. Hodgson, Kerry E. Kilpatrick, Ira M. Longini; March 1977, page 69.

0904 Manufacturing

"An Example of Simulation to Improve Plant Productivity," Walton Hancock, Richard Dissen, Alan Merten; March 1977, page 2.

"The Development of an Automated Program for the Optimal Design of Bulk Cushioning Systems," Don McDaniel; March 1977, page 82

0906 Others

"Optimal Scheduling for Even-Numbered Team Athletic Conferences," Bryan C. Ball, Dennis B. Webster; June 1977, page 161.

1000 ***Management***

1003 Decision Making

"Optimal Scheduling for Even-Numbered Team Athletic Conferences," Bryan C. Ball, Dennis B. Webster; June 1977, page 161.

"Sensitivity of Optimum Workforce Size to Employee Absenteeism in Stochastic Workload Situations," Phiroze J. Nagarvala, Shlomo Globerson; December 1977, page 44.

1107 Project Management

"A Multi-Resource Project Scheduling Algorithm," E. M. Dar-El, Yigal Tur; March 1977, page 44.

1100 ***Material Handling***

1102 Conveyors

"A Model of a Closed-Loop Conveyor with Random Material Flow," Eginhard J. Muth; December 1977, page 345.

1105 Dispatching, Scheduling

"On the Use of A Vehicle Routing Algorithm for the Parallel Processor Problem with Sequence Dependent Changeover Costs," R. G. Parker, R. H. Deane, R. A. Holmes; June 1977, page 155.

"Evaluating a Sequential Vehicle Routing Algorithm," Bruce L. Golden; June 1977, page 204.

1108 Packaging

"The Development of an Automated Program for the Optimal Design of Bulk Cushioning Systems," Don McDaniel; March 1977, page 82.

1111 Storage, Warehousing

"Analysis of the Effect of Buffer Storage Capacity in Transfer Line Systems," Kenjiro Okamura, Hajime Yamashina; June 1977, page 127.

"The Output of A Two-Stage System with Unreliable Machines and Limited Storage," Edward Ignall, Alvin Silver; June 1977, page 183.

"Order Quantity, Product Popularity, and The Location of Stock in a Warehouse," Hoyt G. Wilson; September 1977, page 230.

"The Segregated Storage Problem: Some Properties and an Effective Heuristic," James R. Evans, Frank H. Cullen; December 1977, page 409.

1200 ***Methods Engineering***

1300 ***Quantitative Methods***

"Computational Experience with a Sequencing Algorithm Adapted to the Tardiness Problem," Kenneth R. Baker; March 1977, page 32.

"A Scheduling Algorithm for the Unbalanced Production Line," Russell E. Jacobson, Randall F. Sadowski; March 1977, page 36.

"Reformulation of Special Ordered Sets for Implicit Enumeration Algorithms with Applications in Nonconvex Separable Programming," Dennis L. Bricker; June 1977, page 195.

"Scheduling on Parallel Machines to Minimize Two Criteria Related to Job Tardiness," Thomas S. Nunnikhoven, Hamilton Emmons; September 1977, page 288.

"Minimizing the Sum of Completion Times of n Jobs over m Machines in a Flowshop—A Branch and Bound Approach," S. P. Bansal; September 1977, page 306.

"A Model of a Closed-Loop Conveyor with Random Material Flow," Eginhard J. Muth; December 1977, page 345.

"Scheduling Production on One Machine with Changeover Costs," Wade C. Driscoll, Hamilton Emmons; December 1977, page 388.

1301 Mathematical Models

"On a Multichannel Queueing System with Ordered Entry and Heterogeneous Servers," Masayuki Matsui, Jiro Fukuta; June 1977, page 209.

"Parallel-Channel, Dual-Resource Constrained Queueing Systems with Heterogeneous Resources," G. L. Hogg, D. T. Phillips, M. J. Maggard; December 1977, page 352

1302 Networks

"An Efficient Procedure for Implementing a Dual Simplex Network Flow Algorithm," Richard V. Helgason, Jeff L. Kennington; March 1977, page 63.

"Some Network Flow Models and Heuristics for Multiproduct Production and Inventory Planning," James R. Evans; March 1977, page 75.

"Parallel-Channel, Dual-Resource Constrained Queueing Systems with Heterogeneous Resources," G.L. Hogg, D. T. Phillips, M. J. Maggard; December 1977, p. 352.

"Network Application in Industry and Government," Fred Glover, Darwin Klingman; December 1977, page 363.

1303 Programming: Linear, Nonlinear

"A Survey of Recent Results for a Class of Facilities Location Problems," L. F. McGinnis; March 1977, page 11.

"An Efficient Procedure for Implementing a Dual Simplex Network Flow Algorithm," Richard V. Helgason, Jeff L. Kennington; March 1977, page 63.

"Scheduling Multispecialty Clinics Using Integer Programming," Thom J. Hodgson, Kerry E. Kilpatrick, Ira M. Longini; March 1977, page 69.

"An Application of Facilities Location Theory to the Design of Forest Harvesting Areas," Dennis P. Dysktra, James L. Riggs; September 1977, page 270.

"The Robustness of the Variable S Inventory Model: Some New Results," Cetin Akturk, James V. Jucker; December 1977, page 382.

"The Segregated Storage Problem: Some Properties and an Effective Heuristic," James R. Evans, Frank H. Cullen; December 1977, page 409.

1305 Simulation, Games

"An Example of Simulation to Improve Plant Productivity," Walton Hancock, Richard Dissen, Alan Merten; March 1977, page 2.

"Centralized Scheduling and Priority Implementation Heuristics for a Dynamic Job Shop Model," Rosser T. Nelson, Charles A. Holloway, Ruby Mei-Lun Wong; March 1977, page 95.

"A Versatile Four Parameter Family of Probability Distributions Suitable for Simulation," Bruce W. Schmeiser, Stuart Jay Deutsch; June 1977, page 176.

"Some Relationships Between Dynamic Shop Operations and the Capacity Rating Methodology," W. E. Wilhelm, R. E. Doty; September 1977, page 223.

"Dual-Phase Simulation—A Mixed Time Frame Systems Analysis Tool," David F. Gibson, Edward L. Mooney; December 1977, page 338.

"Sensitivity of Optimum Workforce Size to Employee Absenteeism in Stochastic Workload Situations," Phiroze J. Nagarvala, Shlomo Globerson; December 1977, page 44.

1306 Applied Probability

"A Versatile Four Parameter Family of Probability Distributions Suitable for Simulation," Bruce W. Schmeiser, Stuart Jay Deutsch; June 1977, page 176.

"The Effects of Inspector Error on the Operation of a c-Chart," Alan L. Dorris; September 1977, page 312.

1307 Experimental Design

"Noise Analysis by iso-db A Mapping, Chester J. Kishel, Edward J. R. Hudec; December 1977, page 377.

"Selecting Bayesian Single Sample Acceptance Plans for Quality Control by Pattern Search," Herbert Moskowitz; December 1977, page 396.

1400 *Production and Inventory Control*

"A Scheduling Algorithm for the Unbalanced Production line," Russell E. Jacobson, Randall F. Sadowski ; March 1977, page 36.

"Some Relationships Between Dynamic Shop Operations and the Capacity Rating Methodology," W. E. Wilhelm, R. E. Doty; September 1977, page 223.

"The Transfer of Domestic-Market Production to a Foreign Site," James V. Jucker; December 1977, page 321.

1401 Distribution Systems

"Analysis of the Effect of Buffer Storage Capacity in Transfer Line Systems," Kenjiro Okamura, Hajime Yamashina; June 1977, page 127.

"Evaluating a Sequential Vehicle Routing Algorithm," Bruce L. Golden; June 1977, page 204.

1402 Forecasting

"Synthesizing Short Run Production and Marketing Decisions," Ulf Peter Welam; March 1977, page 53.

1403 Inventory Control

"An Order-Level Lot-Size Inventory Model for Deteriorating Items," Y. K. Shah; March 1977, page 108.

"Order Quantity, Product Popularity, and the Location of Stock in a Warehouse," Hoyt G. Wilson; September 1977, page 230.

"The Robustness of the Variable S Inventory Model: Some New Results," Cetin Akturk, James V. Jucker; December 1977, page 382.

1405 Production Control

"Computational Experience with a Sequencing Algorithm Adapted to the Tardiness Problem," Kenneth R. Baker; March 1977, page 32.

1406 Production Planning

"The Output of A Two-Stage System with Unreliable Machines and Limited Storage," Edward Ignall, Alvin Silver; June 1977, page 183.

1407 Production Scheduling

"An Improved Algorithm for Scheduling Jobs on Identical Machines," J. Wesley Barnes, J. J. Brennan; March 1977, page 25.

"A Multi-Resource Project Scheduling Algorithm," E. M. Dar-El, Yigal Tur; March 1977, page 44.

"Synthesizing Short Run Production and Marketing Decisions," Ulf Peter Welam; March 1977, page 53.

"Some Network Flow Models and Heuristics for Multiproduct Production and Inventory Planning," James R. Evans; March 1977, page 75.

"Centralized Scheduling and Priority Implementation Heuristics for a Dynamic Job Shop Model," Rosser T. Nelson, Charles A. Holloway, Ruby Mei-Lun Wong; March 1977, page 95.

"Scheduling on Parallel Machines to Minimize Two Criteria Related to Job Tardiness," Thomas S. Nunnikhoven, Hamilton Emmons; September 1977, page 288.

"Maintenance Scheduling for Multicomponent Equipment," Roger C. Vergin, Michael Scriabin; September 1977, page 297.

"Minimizing the Sum of Completion Times of n Jobs over m Machines in a Flowshop—A Branch and Bound Approach," S. P. Bansal; September 1977, page 306.

"Scheduling Rules for Parallel Processors," Louis A. Martin-Vega, H. Donald Ratliff; December 1977, page 330.

"Scheduling Production on One Machine with Changeover Costs," Wade C. Driscoll, Hamilton Emmons; December 1977, page 388.

1500 *Quality Control*

1501 Inspection

"A Comparison of Two Optimal Inspection Procedures in the Presence of Errors," Matthew Rosenshine; March 1977, page 19.

"Inspection Policies for A System Which Is Inoperative During Inspection Periods," Hanan Luss; June 1977, page 189.

1503 Reliability

"Availability of Maintained Systems: A State-of-the-

Art Survey," C. H. Lie, C. L. Hwang, F. A. Tillman; September 1977, page 247.

1504 Sampling

"Price Adjusted Single Sampling with Quadratic Indifference," Joseph W. Foster, Oscar T. Perry; March 1977, page 88.

"The Effects of Inspector Error on the Operation of a c-Chart," Alan L. Dorris; September 1977, page 312.

"Selecting Bayesian Single Sample Acceptance Plans for Quality Control by Pattern Search," Herbert Moskowitz; December 1977, page 396.

1600 ***Work Measurement***

1700 ***Data Processing Systems Design***

"Reformualtion of Special Ordered Sets for Implicit Enumeration Algorithms, with Applications in Nonconvex Separable Programming," Dennis L. Bricker; June 1977, page 195.

1703 Analysis Techniques

"Selecting Bayesian Single Sample Acceptance Plans for Quality Control by Pattern Search," Herbert Moskowitz; December 1977, page 396

American Institute of Industrial Engineers, Inc., March 1977, Volume 9, Number 1

AIIE TRANSACTIONS

An Example of Simulation to Improve Plant Productivity	WALTON HANCOCK RICHARD DISSEN ALAN MERTEN	2
A Survey of Recent Results for a Class of Facilities Location Problems	L. F. MCGINNIS	11
A Comparison of Two Optimal Inspection Procedures in the Presence of Errors	MATTHEW ROSENSHINE	19
An Improved Algorithm for Scheduling Jobs on Identical Machines	J. WESLEY BARNES J. J. BRENNAN	25
Computational Experience with a Sequencing Algorithm Adapted to the Tardiness Problem	KENNETH R. BAKER	32
A Scheduling Algorithm for the Unbalanced Production Line	RUSSELL E. JACOBSON RANDALL F. SADOWSKI	36
A Multi-Resource Project Scheduling Algorithm	E. M. DAR-EL YIGAL TUR	44
Synthesizing Short Run Production and Marketing Decisions	ULF PETER WELAM	53
An Efficient Procedure for Implementing a Dual Simplex Network Flow Algorithm	RICHARD V. HELGASON JEFF. L. KENNINGTON	63
Scheduling Multispecialty Clinics Using Integer Programming	THOM J. HODGSON KERRY E. KILPATRICK IRA M. LONGINI	69
Some Network Flow Models and Heuristics for Multiproduct Production and Inventory Planning	JAMES R. EVANS	75
The Development of an Automated Program for the Optimal Design of Bulk Cushioning Systems	DON MCDANIEL	82
Price Adjusted Single Sampling with Quadratic Indifference	JOSEPH W. FOSTER OSCAR T. PERRY	88
Centralized Scheduling and Priority Implementation Heuristics for a Dynamic Job Shop Model	ROSSER T. NELSON CHARLES A. HOLLOWAY RUBY MEI-LUN WONG	95
Extensions of the Multi-Period Facility Phase-Out Model: New Procedures and Application to a Phase-In/Phase-Out Problem	GARY M. ROODMAN LEROY B. SCHWARZ	103
An Order-Level Lot-Size Inventory Model for Deteriorating Items	Y. K. SHAH	108
Book Reviews		113
Research Abstracts	OWEN W. MILLER (Editor)	115
Letters to the Editor		120

INDUSTRIAL ENGINEERING RESEARCH AND DEVELOPMENT

March 1977, Volume 9, No. 1

AIIE TRANSACTIONS

Editorial Staff: JIM F. WOLBRINK, Editor-In-Chief AIIE

RALPH L. DISNEY, Senior Editor
University of Michigan

RICHARD L. FRANCIS, Senior Editor
University of Florida

Publications Policy Board: MITCHELL FEIN, P.E., Chairman
Professional Engineer

ROGER P. DENNEY, JR., P.E.
Medicus Systems Corporation

BENNIE J. BUTTS
Burlington Industries

WILLIAM T. MORRIS, P.E.
The Ohio State University

Editorial Board

S. KEITH ADAMS
Iowa State University

NICHOLAS J. AQUILANO
University of Arizona

JAMES E. BAILEY
Arizona State University

J. WESLEY BARNES
University of Texas at Austin

CHARLES S. BEIGHTLER
University of Texas at Austin

G. K. BENNETT
University of South Florida

RICHARD H. BERNHARD
North Carolina State University

WILLIAM L. BERRY
Indiana University

JOHN E. BIEGEL
Syracuse University

GORDON H. BRADLEY
Naval Postgraduate School

LARRY G. BROWN
Mississippi State University

KENNETH E. CASE
Oklahoma State University

JAMES A. CHISMAN
Clemson University

ROBERT B. CONRAD
University of North Carolina at Charlotte

ROBERT M. CRISP, JR.
University of Arkansas

DOUGLAS C. CROCKER
Eastman Kodak Company

P. M. DEARING
Cornell University

MARK J. EISNER
Exxon Corporation

SALAH E. ELMAGHRABY
North Carolina State University

B. L. FOOTE
University of Oklahoma

JAY GOLDMAN
University of Missouri

RONALD L. GUE
CSF, Ltd.

WALTON M. HANCOCK
University of Michigan

R. J. HERBOLD
Procter & Gamble Company

ALBERT G. HOLZMAN
University of Pittsburgh

ARTHUR P. HURTER, JR.
Northwestern University

JAMES V. JUCKER
Stanford University

KERRY E. KILPATRICK
University of Florida

STEPHAN KONZ
Kansas State University

K. E. LARSON
Western Electric

HERCHEL E. LYNCH
Red River Army Depot

KEITH L. McROBERTS
Iowa State University

WILBUR L. MEIER, JR.
Purdue University

ROBERT W. METZGER
Dow Chemical Company

ARNE MJOSUND
University of Alabama

WILLIAM C. MOOR
Arizona State University

RONALD S. MORRIS
Texas A&M University

RICHARD E. NANCE
Virginia Polytechnic Institute and State University

J. B. NEUHARDT
Ohio State University

BENJAMIN W. NIEBEL
Pennsylvania State University

PETER V. NORDEN
Columbia University
IBM Corporation

JAMES A. PARSONS
Lederle Laboratories

C. C. PEGELS
State University of New York at Buffalo

D. T. PHILLIPS
Texas A&M University

SCOTT T. POAGE
University of Houston

MATTHEW ROSENSHINE
Pennsylvania State University

J. W. SCHMIDT
Virginia Polytechnic Institute and State University

F. STAN SETTLES
AiResearch Manufacturing Company

ROBERT E. SHANNON
The University of Alabama in Huntsville

C. M. SHETTY
Georgia Institute of Technology

WILLIAM E. SOUDER
University of Pittsburgh

WARREN H. THOMAS
State University of New York at Buffalo

GERALD J. THUESEN
Georgia Institute of Technology

RICHARD E. WENDELL
Carnegie-Mellon University

JOHN A. WHITE
Georgia Institute of Technology

RICHARD C. WILSON
The University of Michigan

ROBERT EUGENE WOOLSEY
Colorado School of Mines

DARELL R. ZERBE
Procter & Gamble Company

■ Published quarterly in March, June, September, and December by the American Institute of Industrial Engineers. Executive and Editorial Offices: 25 Technology Park, Norcross, Georgia 30092.

Copyright 1977 by the American Institute of Industrial Engineers, Inc. Reprint permission available from the Editorial Offices.

AIIE accepts no responsibility in connection with any liability which might develop as a result of material published, for opinions expressed are those of the authors and do not necessarily represent the Institute.

Subscription rates: AIIE members subscription information is available upon request: non-members, $35.00 per year (domestic) and $37.00 per year (foreign). Single copy: $10.00. AIIE, 25 Technology Park, Norcross, Georgia 30092.

All issues of *AIIE Transactions* are available from AIIE or on microfilm or Xerox copy from University Microfilms, 300 North Zeeb Road, Ann Arbor, Michigan 48106.

Second class postage paid at Norcross, Georgia and additional mailing offices.

American Institute of Industrial Engineers

President
JOHN F. SWEERS, P.E.
Western Publishing Company

President-Elect and Chief Financial Officer
JOSEPH H. KEHLBECK, P.E.
General Electric Company

Immediate Past President
WILLIAM A. SMITH, JR., P.E.
North Carolina State University

Executive Vice President, Chapter Operations
CLARENCE DARROW, JR.
Federal-Mogul Corporation

Executive Vice President, Divisions, Education and Publications
MARVIN E. MUNDEL, P.E.
M. E. Mundel & Associates

Executive Director
JACK F. JERICHO, P.E.
AIIE

Vice President, International Operations
NELSON C. WALL, P.E.
Georgia Institute of Technology

Vice President, Publications
MITCHELL FEIN, P.E.
Professional Engineer

Vice President, Education and Professional Development
FRANK A. TILLMAN, P.E.
Kansas State University

Vice President, Industry and Management Divisions
A. JOHN GEIS, P.E.
Bright of America, Inc.

Vice President, Systems and Technical Division
WOLTER J. FABRYCKY, P.E.
Virginia Polytechnic Institute and State University

Vice President, Region I
William C. Smith
Pratt & Whitney Aircraft

Vice President, Region II
GARY E. WHITEHOUSE, P.E.
Lehigh University

Vice President, Region III
GRADY F. PETERSON, P.E.
Western Electric Company

Vice President, Region IV
GEORGE F. SCHRADER, P.E.
Florida Technological University

Vice President, Region V
JOHN J. GEIER
Eastman Kodak Company

Vice President, Region VI
FREDERICK G. KING, P.E.
Youngstown Sheet and Tube Company

Vice President, Region VII
SIDNEY G. GILBREATH, III, P.E.
Tennessee Technological University

Vice President, Region VIII
PAUL F. DAVIS
United Airlines

Vice President, Region IX
FRANCE A. MEIER
University of Texas

Vice President, Region X
JOSEPH F. RODGERS, P.E.
Raychem Corporation

Vice President, Region XI
FRANK J. MILES
Maytag Company

Vice President, Region XII
ROBERT J. VERDERBER, P.E.
Consulting Industrial Engineer

Vice President, Mexico (Region XIII)
JAVIAR R. ARIAS SAN ROMAN
Ing. Javiar Arias Y. Associadas

Vice President, Canada (Region XIV)
JOHN C. COCHRANE
The Steel Company of Canada

Department Editors

Applied Optimization	Professor Gordon H. Bradley Department of Operations Research & Administrative Sciences Naval Postgraduate School Monterey, CA 93940
Applied Probability and Statistics	Dr. Darell R. Zerbe Winton Hill Technical Center The Procter and Gamble Co. 6071 Center Hill Road Cincinnati, OH 45224
Facilities Design/Location Analysis	Professor John A. White School of Industrial & Systems Engineering Georgia Institute of Technology Atlanta, GA 30332
Manufacturing and Materials	Professor John E. Biegel Department of Industrial Engineering & Operations Research Room 441, Link Hall Syracuse University Syracuse, NY 13210
Organization Planning/Engineering Economics	Professor Gerald J. Thuesen School of Industrial & Systems Engineering Georgia Institute of Technology Atlanta, GA 30332
Scheduling, Planning and Control	Professor Salah E. Elmaghraby Industrial Engineering Department North Carolina State University Raleigh, NC 27607
Simulation and Interactive Games	Professor Richard E. Nance Department of Computer Science Virginia Polytechnic Institute & State University Blacksburg, VA 24061
Work Measurement/Methods/Ergonomics	Professor Walton M. Hancock The University of Michigan 109 Observatory Ann Arbor, MI 48109

Instructions for Authors

Brief information for contributors is given below. Please see complete instructions inside back cover.

Four copies of all manuscripts should be sent directly to:

Editor-In-Chief
AIIE TRANSACTIONS
25 Technology Park
Norcross, GA 30092

Failure to submit the necessary four copies will result in delays in the reviewing process.

Where possible, the author is encouraged to designate, in the letter of transmittal, the proper Department to review the paper. It is certainly recognized that some manuscripts may not fall clearly in any of the above departments. In such cases it is desirable for the author(s) to so designate. The Editor-In-Chief will make the assignment to the department and so inform the author(s) in the adknowledgement card, sent upon receipt of the manuscript.

Manuscripts intended for the Development and Applications Award consideration should also be submitted to the Editor-In-Chief.

An Example of Simulation to Improve Plant Productivity

WALTON HANCOCK
SENIOR MEMBER, AIIE
RICHARD DISSEN
ALAN MERTEN
The University of Michigan
Ann Arbor, Michigan 48109

Abstract: In an effort to increase the productivity of a small parts stamping plant, a computer simulation was used to determine the relative economic advantages of changing the crewing, increasing the spare tooling, or improving the die preparation. The percent utilization and the cost per 1,000 parts produced were chosen as the measures of performance. Problems associated with data collection are discussed. The output of the simulation and the results that were achieved with a demonstration group of eight machines are presented. Machine utilization was increased from 37% to 60% with a potential increase to 76%. The appropriate management information system to maintain high productivity is discussed.

■ This paper is intended to describe an actual situation where simulation was used to substantially improve productivity. Since the actual data are confidential, fictitious data are used; however, the procedures and the relative results should be of interest to professionals involved in improving productivity. Other published examples of the use of simulation to improve productivity are rare. References [1], [2], and [3] are examples of similar uses of simulation where actual results are reported.

The company involved has a number of manufacturing plants located throughout the world, and its principal products are stamped parts from brass and aluminum. Each plant has from 20 to 50 stamping presses whose stampings vary in complexity and volume according to demand. The dies used are multi-staged progressive dies designed and manufactured by the parent company. Most of the dies can produce more than one part.

The top management of the company was aware that the productivity of the different plants varied widely; however, the firm was growing rapidly and had not taken time to make detailed comparisons of the differences between plants. The percentage of machine utilization of the plants ranged from 20% to 85% and is defined as:

$$\frac{(\text{number of parts produced} \div \text{standard running speed for the part}) \times 100}{\text{number of hours in run mode}}$$

Note that the utilization excluded set-up time, but did not exclude all other factors that might cause the machines to be down or to run slower than the rated speeds.

Received August 1975; revised December 1975; revised August 1976.

Investigating Committee

In order to learn as much as possible concerning those factors that affected productivity, a plant near the company's main headquarters was chosen as an experimental site. This plant had an average utilization of 38% and, at the time the project started, had been run at six-day capacity for several months. A committee was established consisting of the Plant, Production, and Maintenance Managers, the Chief Die Engineer, two Staff Industrial Engineers, and a consultant who chaired the committee and was asked to communicate periodically with the top management of the company.

At the first meeting of the committee, the problem of low utilization was discussed and the following list of possible reasons was made:

1. Poor equipment maintenance
2. Not enough machine operators
3. Inadequacy of spare tooling (the dies have precision machined inserts which wear and break)

4. Run lengths too short
5. Inadequate die preparation

When the committee was asked to list the reasons in order of importance, no agreement could be reached. The chairman, in an effort to get some rank ordering, attempted to secure an estimate of the increase in percent utilization that was likely to occur for each item listed. Again, no conclusions were reached and very few estimates were given. The possible reason for the lack of response was that the method of approach was new to the group.

At the second meeting, a new approach was tried. The chairman asked the question "Suppose the equipment was put in first-class condition, the crews were properly established, spare tooling was always available, the dies issued to the pressroom were to specification, and the product mix was the same as it is now, what would be your estimate of the utilization you could achieve?" Again, there was considerable discussion, but this time the members of the committee were able to provide an overall utilization figure of 60%. When they were asked to partition the increase of 22% among the contributing items, no response was forthcoming; the members of the committee had never studied the individual components.

The committee then was asked to give estimates of the cost to provide this "ideal" working environment. The results were as follows:

1. Rebuild equipment–6 man-years of tool maker time plus 2 additional machine tools
2. Machine operators–6 additional machine operators for the plant
3. Spare tooling–$300,000 or the addition of 5 more tool and die makers for a two-year period
4. Die preparation–4 additional tool and die makers to be added to the die staff

The committee prepared a report providing costs for the above items and comparing them to the increase in productivity. The comparisons were quite favorable, indicating the company would recoup the entire cost in eight months and the continuing costs of additional labor in three months; however, there was no solid evidence that the committee's estimate of 60% utilization was attainable.

Experiment

To provide more solid evidence, eight presses, the number normally assigned to a crew, were established as an "island group," i.e., a group of presses independent from the remainder of the plant. The committee established the size of the island group at eight presses because the machine operators had difficulty monitoring more than eight machines due to the physical size of the presses. In addition, it was felt necessary to hold the die repairmen responsible for a specific number of presses in order to insure complete familiarity with the machines. These "island group" presses had the following characteristics:

1. All eight machines were less than one year old–they were in new condition.
2. The crew, which consisted of Machine Operators, Die Setters, Die Repairmen, and a Group Leader, was assigned as stipulated by the committee.
3. Twenty-five dies were selected–the parts made for these dies had an average utilization of 38% on the presses of the same design as those selected for the "island group."
4. Operating instructions were issued stating all dies run in the "island group" would be machined to specifications before being issued to the pressroom.
5. Instructions were issued to make spare tooling available at quantities specified by the spare tooling inventory system when the die was issued to the pressroom. The quantities of spare tooling previously had been allowed to go substantially below the specified inventory levels.

Several of the above conditions were found to be much harder to attain than others. The most difficult were establishing the spare tooling inventory and machining the dies to their specifications, since both of these requirements depended on the die drawings being current. Meeting these two requirements took one and a half months. The lag time on procuring spare tooling was three to six months, making it very difficult to establish the ideal situation quickly.

The utilization percentage began to increase steadily over time as indicated in Fig. 1. After ten weeks, 50% utilization

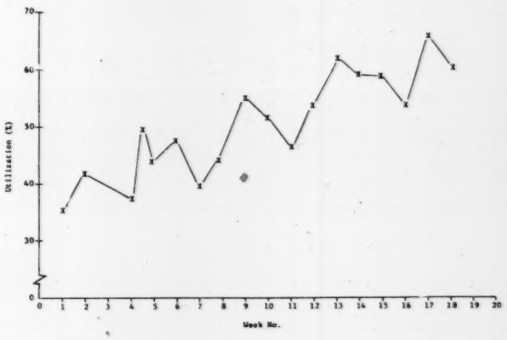


Fig. 1. Utilization (%) vs. Week Number for the Island Group.

was achieved; however, no member of the committee understood why and there was a feeling that the results would not last. Nor was the variation from week to week in Fig. 1 understood. There were predictions that the results achieved were due to certain so-called "dog" dies not being run as often as the other dies. In addition, the committee was repeatedly asked by management what impact the changes in sewing had on the results, and if the addition of more operators would improve utilization even more.

The committee decided to develop a computer simulation of the "island group" in an attempt to gain more insight as quickly as possible.

Data Reduction and Its Implications

The writing of the simulation sharply defined what data would be needed. Without going into detail, it was discovered that although most of the data could be obtained, lengthy data reductions were required to get the data into the proper form. The data, obtained after two man months of effort by the Staff Industrial Engineers, were as follows:

1. Time at which set-ups occurred and duration of set-up
2. Number of parts per running hour and duration of job
3. Time at which die repair was needed and duration of repair
4. Time at which loading (and unloading) of material occurred and duration of the loading (and unloading)
5. A decision matrix of which tasks each labor class performed and their task priorities
6. The actual percentage utilization over the period in which the data were collected

The reduction of the above data provided considerable insight into the more important factors affecting the utilization. For example:

1. The die failure rate had a mean of 0.37 failures/machine running hour with a range from 0.15 to 0.78. Subsequent references to "die failure rate" will be in terms of mean number of failures per machine running hour. The average failure rate per week was plotted against the percentage utilization/week as shown in Fig. 2. The results are quite clear: As the die failure rate improves, the utilization improves.

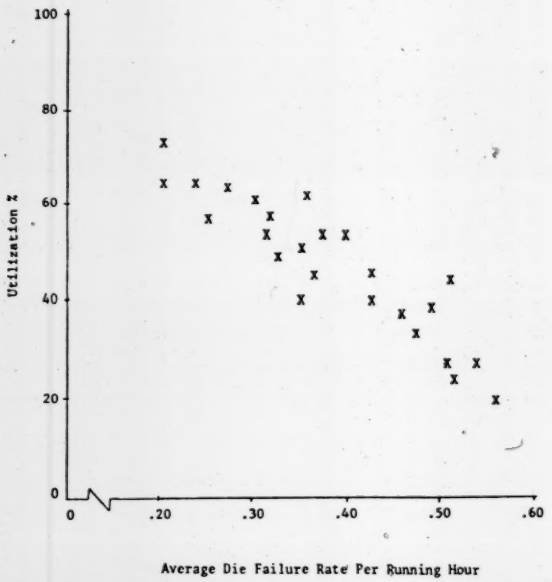


Fig. 2. A plot of utilization % per week vs. average die failure rate.

2. The average die repair time was 1.0 hours, and 90% of the repair times were between 0.5 hours and 1.5 hours. This consistency of repair times was investigated and it was discovered that most of the repair time was consumed opening and closing the die. Tool replacement and adjustment took most of the rest of the time. The main conclusion was that the repair time did not vary substantially with who performed the repair. This was evidence that the training of some of the people was not a major source of poor utilization.
3. Each die would run several part numbers and each part number had its own failure rate. Further investigation revealed that the die preparation personnel did not always know what part numbers were going to be run until after the die was in the press. Frequently, certain stations which were for specific part numbers were not adequately prepared, and the die failure rates reflected this situation.

These findings led to increased attention to the preparation of the dies. The committee agreed that any part number that had a failure rate of above 0.45 would be examined and, wherever possible, instructions issued to make improvements. This process revealed that of the eight dies in this category all but one had major design deficiencies. Most of these dies had been designed several years prior to this study and therefore did not take advantage of recent technological advances. Several dies required two or three cycles through the review process in order to bring about a more acceptable failure rate. Substantial costs were incurred, in several cases exceeding budget limits, but when higher management was made aware of the potential savings obtainable by improving the failure rates, permission was readily obtained to exceed die repair budgets.

Computer Simulation

The result of these efforts was that the utilization rate stabilized at approximately 60%—the intended goal. However, the question concerning the best crew size and configuration still remained. Also, the committee began to reconsider the 60% and concluded that 70%-80% was attainable. With these factors in mind, the simulation was written in IBM's General Purpose System Simulator (GPSS). The actual distributions of job sequence, run length, set-up times, and wait for tools times were used in all cases. Job sequence and run length were randomly selected from these distributions initially, but the same sequences were then used on repeated runs with different crewing configurations.

One of the more difficult aspects of the simulation was to determine what the various members of the press crews actually did. The reason this was so difficult was that when the project was initiated, four classes of people were working on the machines. Each class had certain responsibilities according to the job evaluation system. However, as the project progressed, the employees redefined their traditional roles so that all personnel, except the machine operator who was the least experienced, would perform die repairs as

Table 1: Jobs performed by each labor category (X). Numbers in parenthesis indicate priorities with the highest priority being (1).

	A Complete set-up	B Changeover in press	C Changeover in shop	D_A Reel change (quantity)	D_B Reel change (type of reel)	E Stock change	F Factory order change	Failure
Machine Operators				X (1)		X (1)		
Die Setters	X (2)	X (2)	X (2)		X (2)	X (2)	X (2)	X (1)
Die Repairmen	X (2)	X (2)	X (2)		X (2)	X (2)	X (2)	X (1)
Combination Die-setter, Die-repairmen	X (2)	X (2)	X (2)		X (2)	X (2)	X (2)	X (1)

their first priority. Table 1 gives the priority and job matrix used in the simulation. The machine operators' first priority jobs (and their only jobs) were the changing of reels of the brass strip stock from which the components were stamped and the removal and inspection of reels of stamped components.

The consultant attempted to provide the committee with near optimum crewing solutions using analytic techniques. He stressed the assumptions he was making concerning the distributions, the problems connected with dealing with the interactions between different types of set-ups and machine failures, and with crew members with specific and overlapping responsibilities. Feller [4] gives the solution for r repairmen servicing m machines. The assumptions that Feller made, that the committee would not accept, were (1) that repair times were exponentially distributed, (2) that all members of the crew did the same work, and (3) that such factors as wait for tools, different types of set-ups, and reel changes could not be taken into account. The committee decided that they would prefer to use simulation so that the assumptions could be kept to an absolute minimum. They also felt that results achieved by simulation would have more credibility with their supervisors because such results could be more easily explained.

The only assumption made in the simulation was that the time between failures of a machine had a negative exponential distribution. This assumption was necessary because the data collection process which was adequate for obtaining the number of failures and the resultant repair times was inadequate for getting the time between failures.

An Example of the Simulation Output

Table 2 is an example of the output format of the simulation. Several important points should be made concerning this output.

1. The word "Replications" in line 1 of the output could also be read as "Weeks."

Table 2: Typical Output of Simulation.

19 Replications Completed
2 Machine Operators. . . . . .$ 700.00
3 Die Repairmen 1,200.00
Factory Overhead 5,000.00
Total $ 6,900.00
Utilization: 51.51%
Cost/1,000 = $0.613

Time spent in each event

Set-up Type	Minutes	Percent of total time
A	402	1.09%
B or C	166	0.45%
D	43	0.12%
E	117	0.32%
F	12	0.03%
Other Events		
Wait for personnel	7,717	20.92%
Reel Change-A	858	2.33%
Wait for tool	2,539	6.88%
Repairing	6,414	17.39%
Stamping	18,617	50.47%
Die failure rate	.37/Hr./Machine	

2. Crewing section—The number of different types of crewmen is posted as well as the associated cost. For each simulation, the crew configuration is an input parameter. Please note that only two personnel categories are listed, machine operators and die repairmen, because the development of Table 1 revealed that die setters and die repairmen did the same jobs.
3. The output measures—The percentage machine utilization and the cost per thousand (cost/1,000) are listed. Utilization, of course, was one of the primary measures of success of the project. However, cost/1,000 was a more widely used term in the company. Also, the real objective was to minimize the cost/1,000 so the parameter was introduced here. It is important to realize that the plant employees were not aware of (and did not seem to have much feel for) the cost/1,000, whereas the percentage utilization was very meaningful and relevant to them. On the other hand, the controller and many higher managers did not relate to utilization as well as they did to cost/1,000.
4. Time spent in each event section—This section gives the data on the amount of time and the percentage of the total time spent in each major event of the production process. Although the project was primarily concerned with those items listed under "Other Events," it is interesting to note the relatively small percentage of time consumed by set-ups of all types—2.01%.
5. Die failure rate—The average die failure rate was specified as an input parameter in each simulation run. The experienced die failure rate was included as part of the simulation output during the first runs. In all cases it was found to be (at 19 weeks) within 0.005 of the input value. This was very satisfactory and was excluded from further output.

6. All data which were output from the simulator are expressed as *cumulative averages*. For example, after 5 weeks (replications), the machine utilization figure was the average utilization over the 5 week period. After 10 weeks, this would be the utilization over the entire 10 week period. Output was only available in 5 week intervals to keep simulator costs low and because preliminary runs revealed that the variation from week to week stablized after seven to nine weeks. The range of the data for 15 and 19 weeks was used as a measure of the "noise" in the system. The range was preferable because
   a. The plant personnel could easily understand the concept of range. Great difficulty was experienced in discussing statistical variance.
   b. Successive data values for any output variable are serially dependent, and thus, a computation of the variance of such output measures was felt to be statistically inappropriate on theoretical grounds.
7. This entire productivity problem should be viewed as a stochastic process. As such, the cumulative average of each output variable will converge to a limit as time increases without bound. Because of the dependence of the measures of the output variables (see item 6), the best indicator of the variability of the simulation at any point in time, i.e., the noise left in the system, was felt to be simply the range of the output variable near that point.
8. A simulation span of 19 weeks was chosen as an appropriate time frame for this study. This is a short enough length so that the time and money spent doing the simulations was not excessive, but also long enough so that the variability of each simulation had been dampened sufficiently and an estimate of the noise range could be made. Of the 26 cases where the data are available at 15 and 19 weeks (only the 19 week output was obtained in 8 cases), the range of cost/1,000 between 15 and 19 weeks in 25 of 26 cases was no more than \$0.026. The average range of cost/1,000 for the 26 cases was \$0.012.
9. In order to calibrate and help validate the simulator, a data base consisting of actual data concerning the intricacies of the entire production process was set up. After the construction of the simulator, this data base was then used as input to the simulator in order to test its validity. The actual machine utilization during the data collection period and the simulated machine utilization differed by only 0.5% while the actual cost/1,000 and the simulated cost/1,000 were within \$0.01 of each other. Such close agreement in these results helped to establish the credibility of the conclusions that resulted from the analysis of the simulation output.

Table 3: Simulation results for cost, utilization, and wait.

Die Failure Rate	Machine Operators	Die Repairmen	Cost per Thousand	Machine Utilization (%)	Personnel Utilization (%)	Wait for Personnel (%)
.17	3	3	.40	76.0	25.9	3.5
.17	1	3	.42	74.0	37.0	5.9
.17	1	5	.42	75.9	25.8	3.7
.17	1	4	.43	76.3	30.1	3.8
.17	1	2	.44	64.9	43.2	17.0
.17	2	2	.44	67.1	33.4	14.9
* .17	2	3	.46	74.7	29.9	5.3
* .17	2	4	.46	78.0	24.8	1.3
.17	3	4	.48	78.8	21.7	0.6
.17	4	3	.48	75.9	21.0	3.5
.17	1	6	.49	76.1	21.4	2.9
.17	3	2	.50	64.8	24.7	17.2
* .25	2	4	.48	70.9	28.9	2.3
* .25	1	4	.49	68.5	34.7	5.4
* .25	2	3	.50	65.3	32.4	10.4
.25	2	5	.50	71.9	25.6	0.8
.25	1	5	.51	70.9	29.3	2.7
.25	2	6	.51	72.8	22.6	0.2
.25	3	4	.53	71.5	24.5	2.1
* .37	1	5	.49	62.8	36.7	3.9
* .37	2	4	.50	61.3	36.0	6.0
* .37	2	5	.52	64.3	33.1	1.3
.37	1	6	.53	63.0	31.3	2.6
* .37	2	3	.61	51.5	33.4	20.9
.37	2	7	.61	64.7	24.8	0.1
.37	3	4	.62	60.8	29.0	7.1
.37	2	6	.62	65.3	28.2	0.4
* .37	3	3	.69	53.0	28.4	19.5
* .56	1	6	.67	54.2	37.4	3.0
.56	2	4	.69	47.5	37.6	15.1
* .56	1	5	.70	52.7	40.7	6.9
.56	2	5	.72	53.8	36.2	5.4
.56	1	7	.72	55.0	33.4	1.6
* .56	2	3	.90	38.4	33.5	31.0

*Original data points at the time of work. The additional points were run and included for publication purposes.

Relationships Between Costs and Die Failure Rate

Table 3 is a summary of 19 week simulations for 34 experimental conditions. The simulations were run at average die failure rates of 0.17, 0.25, 0.37, and 0.56 per machine/hour. Different crewing configurations were run at these rates in an attempt to obtain the lowest cost/1,000. The main conclusions that can be drawn from this summary are as follows:

1. Given optimal crewing, minimum cost/1,000 is dependent upon the die failure rate.
2. Given a fixed die failure rate, minimum cost/1,000 is dependent upon the crew configuration.
3. Minimum cost crewing for simulation points examined, using cost/1,000 as a measure, is dependent on the die failure rate.

For example, minimum cost crewing changes as follows:

Die Failure Rate	Crewing at Minimum Cost/1,000		Minimum Cost/1,000
	Machine operator	Die repairman	
.17	3	3	$0.40
.25	2	4	0.48
.37	1	5	0.49
.56	1	6	0.67

In order to more clearly display the effect of crewing on cost/1,000, Table 3 and Figs. 3, 4, 5 and 6 are included. In examining these figures, please note that the minimum cost/1,000 differences are less than the

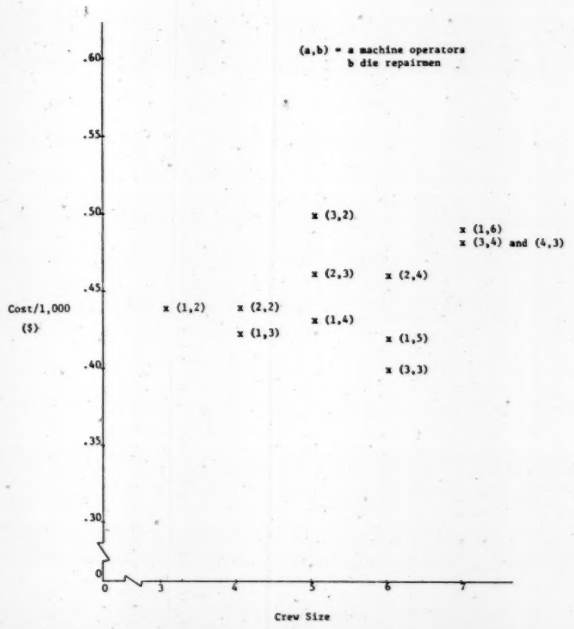


Fig. 3. Plot of cost/1,000 versus crew size. Die failure rate = .17.

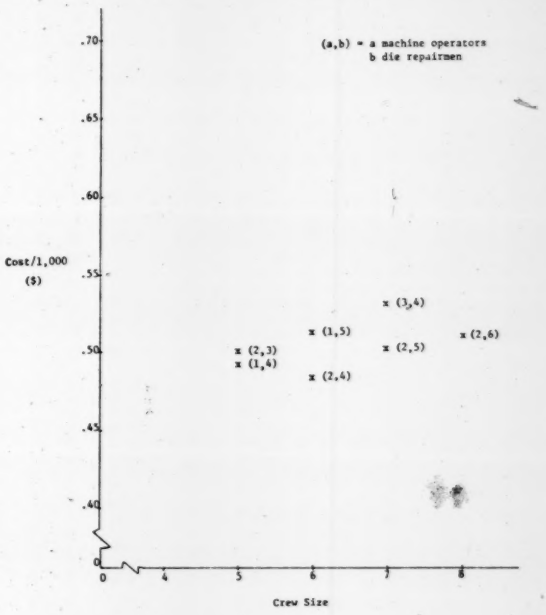


Fig. 4. Plot of cost/1,000 versus crew size. Die failure rate = .25.

noise range of the simulation, $0.026, and thus may be due only to simulation error. The plant personnel reacted to this point by requesting the lowest cost

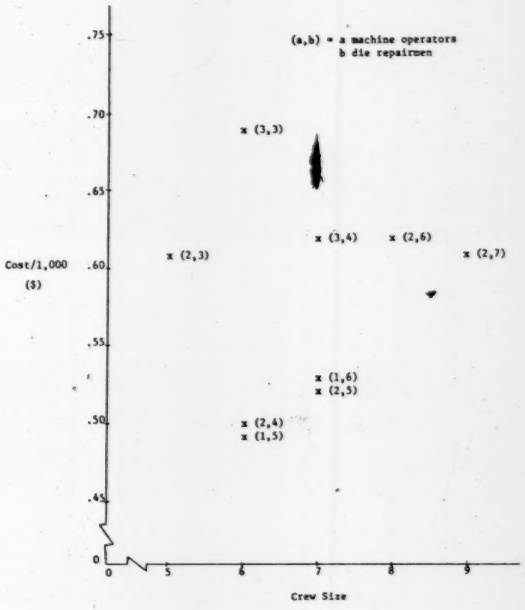


Fig. 5. Plot of cost/1,000 versus crew size. Die failure rate = .37.

simulated output, because of the need to have definite recommendations. The differences between the minimum and second from minimum cost/1,000 are as follows:

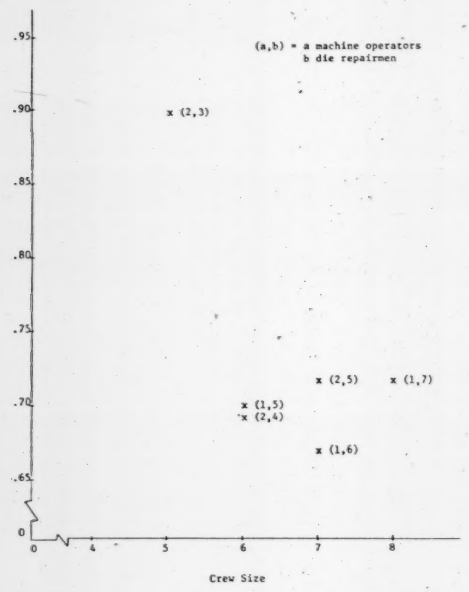


Fig. 6. Plot of cost/1,000 versus crew size . Die failure rate = .56.

Figure Number	Die Failure Rate	Minimum Cost/1,000	Crewing at Minimum Cost/1,000		Second from Best Cost/1,000	Crewing at Second from Best Cost/1,000		Difference in Cost/1,000
			Machine Operator	Die Repairman		Machine Operator	Die Repairman	
3	.17	$.40	3	3	.42	1	5	.02
4	.25	.48	2	4	.49	1	4	.01
5	.37	.49	1	5	.50	2	4	.01
6	.56	.67	1	6	.69	2	4	.02

4. The machine utilization percentage is dependent upon the die failure rate if optimal crewing is assumed.
5. For a fixed die failure rate and changing crew combinations, machine utilization is generally uncorrelated with cost/1,000.
6. Since the "island group" had a mutual die failure rate of 0.37/machine/hour, the minimum cost crewing would give them an average of 63% utilization with a cost/1,000 of $0.49. Lower cost would require reduction in the die failure rate.
7. The job assignments (Table 1) require that at least one machine operator be part of the crew. For higher die failure rates (Figs. 5 & 6), the minimum cost/1,000 was with one machine operator. If the die repairmen were permitted to do *all* the tasks of the machine operators, then it might be possible to reduce costs even further.

Relationship Between Wait for Personnel and Die Failure Rate

Table 3 also summarizes wait for personnel (%) vs. die failure rate. In all cases the minimum cost crewing was where the machine waiting time was less than 4%. The economic impact of low machine waiting time is understandable when one examines the ratio of fixed to variable costs. Since the overhead is approximately 2½ times the labor cost, the minimum cost/1,000 will have low machine wait times. The optimal wait times would occur where the marginal cost of adding crew is balanced by the marginal savings achieved from reduced machine idleness. If the costs change, then the best crew choice may also change.

Relationship Between Utilization of Crews and Die Failure Rate

Table 3 shows the amount of utilization of the crews for all reasons vs. the die failure rates. Note that the minimum cost solutions exist when the crews are utilized between 25-38%. Therefore, the minimum cost solutions are where the crew is *idle* 62-75% of the time. This statement caused more discussion among the plant personnel and particularly the Staff Industrial Engineers than any other aspect of the simulation results. The idea that machine crews should be allowed to be idle 62-75% of the time was very difficult to accept. When it was pointed out repeatedly that the business of the crews was to be available to keep the presses running, the concept was a little easier to accept. The idea of having 62-75% "excess" labor was also very difficult for the company controller as well as the division managers to accept. This is not an isolated result however, and similar results occur in service functions where random demands are placed upon people or facilities.

Relationship Between Wait for Tooling and Die Failure Rate

Figure 7 is a plot of the wait for tool (%) for the optimum crewing vs. the die failure rate. The wait for tool category is a portrayal of the percentage of time spare tooling would not be available if the conditions that existed during the data collection period were maintained. Figure 7 shows that as the failure rate increases the wait for tool (%) also increases. Spare tooling inventories can be much lower with lower die failure rates. The data also provide a basis for doing a cost analysis to see if the annual spare tooling inventory carrying cost of $30,000 would be justified in terms of the increase in productivity (assuming the $300,000 have been spent to bring the inventory to acceptable levels). For example, at a die failure rate of 0.37, 6.88% of the

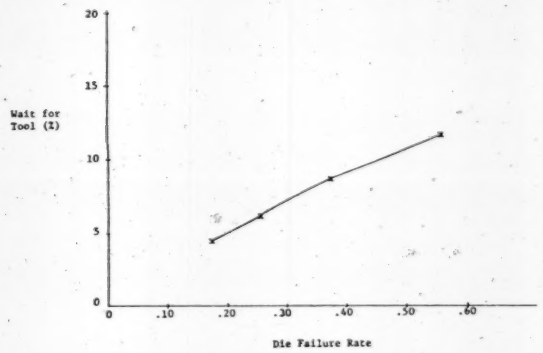


Fig. 7. The wait for tool (%) for the minimum cost crewing versus die failure rate.

time was spent waiting for tools (see Table 2). If this 6.88% were reduced to 0.00% by the means of carrying a larger spare tooling inventory, then the cost/1,000 in Table 2 would be decreased from \$0.61/1,000 to approximately \$0.57/1,000,* the savings could be \$483/week, and it would take approximately 62 weeks to recover the additional inventory cost of \$30,000 per year. Therefore, reducing wait for tool time to zero is presently economically unsound, though it may be a defensible strategy if maximization of production is the primary objective. Of course, if the die failure rate was 0.25, then the carrying cost would be lower, thereby reducing the number of weeks to recover the additional costs.

The Effects of Weekly Variation and the Critical Control Parameters

Figure 8 shows the weekly results of the conditions portrayed in Table 2. Note the variation from week to week in both the percent machine utilization and the pieces produced. This wide variation under steady state conditions for crewing and die failure rate has an important impact on the choice of management control systems. Weekly reports of utilization and/or output are relatively meaningless unless control chart techniques are used. However, even if the control chart techniques are used, an out-of-control figure will not be very helpful because no one will know with assurance what to do unless more basic information is available. Since the weekly figures are the results of the performance or lack of performance of other critical events, the control systems ought to be directly based on the more basic events as follows:

1. The die failure rate—Plant personnel should record the die failure rate for each production run on each die. This is relatively easy to do because the failure rate is merely:

   the number of failures incurred while producing part number x/the number of hours the machines were in run mode for part number x

   For example, if a die fails 10 times in 20 hours of run time while running part x, the die failure rate is $10/20 = 0.5$. The die preparation department should be responsible for keeping the die failure rate of all dies below a certain number established by policy. This index will provide a measure of performance of the die preparation department, a measure which was not previously available but which is fundamental to minimum cost production.
2. Optimum crewing for the average failure rate—A policy decision concerning the average machine failure rate should be established and reviewed periodically. The average machine failure rate was 0.56 at the beginning of the island experiment and was decreased to 0.25 at the end of four months. A figure of 0.25 was established as the maximum die failure rate acceptable to production personnel. This figure was determined on the basis of the ability of the plant engineering staff to improve die design and the capability of outside resources to supply spare tooling. Any time a die failed to perform at this rate, the die preparation department was to be held responsible and accountable. The crewing was established at the 0.25 figure. A crew of two machine operators and four die repairmen, according to Table 3, are capable of attaining 70.9% utilization. If the spare tooling is established to cut the wait for tooling time to zero, then the utilization that the "island group" can attain will be approximately 75.8%.

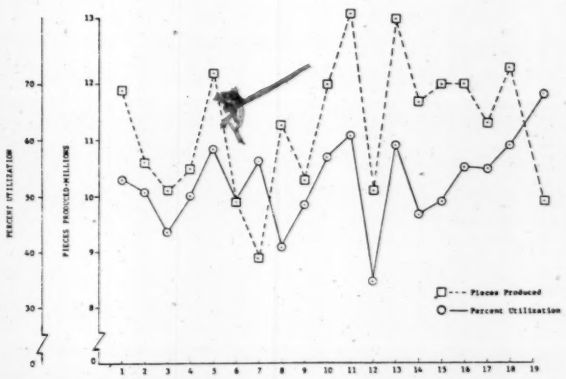


Fig. 8. Results by week for the overall output in Table 2.

Summary and Conclusions

The establishment of an acceptable die repair rate, proper crewing, and satisfactory wait for tooling provides the plant operating people with a basis for measuring three

*Reduction in waiting for tools would cause the number of minutes incurred in the remaining events to increase slightly. This has been approximated by prorating in the calculations.

critical aspects of their operations. If they keep the die failure rates under control, staff the presses correctly, and establish proper spare tool inventories, these control procedures will cause the weekly percentage utilization figure and the pieces produced per week to be within the range of random variation around a very acceptable and predictable mean (75.8%). The plant people will now know what to do in the event production decreases if they use the primary measures correctly. In contrast, when the project started, they did not know what to do when the percentage utilization and pieces produced were the primary control data. Likewise, the controls of die failure rates, the correct staffing of presses, and proper spare tool inventories will result in costs/1,000 which will be at near minimum. The cost/1,000, of course, was the most important criterion, but had to be translated into die failure rates, crewing configurations, and spare tool inventories for use by plant personnel.

During the course of the experiment, machine utilization was increased from 38% to 60%. As 60% was achieved, the committee's knowledge about the critical aspects of the operations had increased to the point that they no longer believed that the potential of the presses had been achieved. The knowledge gained through the development and use of the simulation was of substantial aid in the understanding of the relative importance of die failure rates, crewing configurations, and die inventories. The need to monitor these aspects of the operations rather than the more classical measures of output per week was fully established.

References

[1] Ashour, Said, and Bindingnavle, S. G., "An Optimal Design of a Soaking-Pit Rolling-Mill System," *Simulation*, **18**, pp. 207-214 (June 1972).

[2] Blood, V. Steven, and Santos, William V., "Simulation Cuts Inventory Center Costs," *Industrial Engineering*, **3** 6, pp. 25-27 (June 1971).

[3] Falugh, R. S., and Delporto, R. W., "Simulation of Automated Material-Handling Systems Using GPSS," *Simulation*, **17**, pp. 65-70 (August 1971).

[4] Feller, William, *An Introduction to Probability Theory and Its Application*, John Wiley and Sons, Inc., Third Edition, (revised printing), **1**, pp. 460-468 (1967).

Dr. Walton M. Hancock is Professor of Industrial and Operations Engineering and Hospital Administration at The University of Michigan. He holds bachelor's, master's and doctor's degrees from The Johns Hopkins University. Dr. Hancock was formerly Chairman of the Department of Industrial and Operations Engineering at The University of Michigan and was Manager of Industrial Engineering and Quality Control for the Lord Baltimore Press. His current research work is in Hospital Scheduling Methodologies, Hospital Bed Sizing Studies and Management Information Systems. A Senior Member of AIIE, Dr. Hancock received the Institute's Work Measurement Award in 1972. He is a Departmental Editor of *AIIE Transactions* and belongs to MTM Association and ASEE.

Dr. Alan G. Merten is Associate Professor of Management Science at The University of Michigan. The Design and Restructuring of Large Data Bases and the Modeling of Computer Systems are his major research interests at the present time. Dr. Merten received his undergraduate degree in Mathematics from the University of Wisconsin, and MS and PhD Degrees in the area of Computer Science from Stanford University and the University of Wisconsin respectively. He is a member of ACM.

Mr. Richard L. Dissen is a Consultant with American Management Systems, Inc., in Oak Brook, Illinois. His responsibilities include the general design, detail design and programming of large-scale systems, often using data base management systems. Mr. Dissen holds the BSE and MBA Degrees from The University of Michigan where he was a research assistant for three years. He was elected to membership in Alpha Pi Mu and Tau Beta Pi and is a member of ACM.

A Survey of Recent Results for a Class of Facilities Location Problems

L. F. McGINNIS
MEMBER, AIIE
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Abstract: With advances in computer technology and the state of the art in integer programming, many previously intractable problems are now being solved. This survey attempts to determine the extent to which new developments have been applied to certain facilities location problems. Some promising new techniques are highlighted.

■ For the purposes of this survey, a *facilities location problem* is defined as the problem of selecting sites at which facilities are to be opened, determining the production at each site and the distribution required to meet known demands at minimum cost. Classic examples are the location of warehouses in a distribution system and the location of plants in a multi-plant production system.

Facilities location problems have received a great deal of attention in the past decade, resulting in a proliferation of models and analyses (see, e.g., [30]). The objective of this paper is to survey a subset of the vast literature on facilities location problems, namely, the subset concerned with the development and testing of mixed-integer programming (MIP) solution procedures for problems in which the sites must be chosen from a given catalogue of available sites. Such a catalogue of sites may be apparent from the circumstances in a particular problem (e.g., only certain sites are suitable because of land costs, environmental considerations, etc.) or may be delimited from a continuum of possible locations on a network to location at the nodes via the results of Hakimi and Maheshwari [25] and Wendell and Hurter [41].

The articles surveyed present two types of results: specialized versions of general purpose MIP algorithms or techniques, and problem specific devices for accelerating the solution procedures. It will be assumed that the reader is familiar with the general purpose algorithms. (Much of the notation and terminology used are based on [15] and [21]).

The survey is organized as follows. First, there is a brief review of MIP models of facilities location problems, and some special structure in the models. Next is a survey of the tactics employed in many enumeration-based algorithms for simple facilities location problems and the reported results. Following that is a discussion of other reported solution procedures, some related problems, and some promising new developments.

To facilitate the discussion to follow, certain conventions will be observed. A problem name for a problem with some 0-1 variables, e.g., P, when written with an over-bar, e.g., $\bar{P}$, will denote the problem with the integer restrictions relaxed to simple upper bounds. Also $v(\cdot)$ will denote the solution value for problem $(\cdot)$, e.g., $v(P)$ is the value obtained by solving problem P.

The Basic Model

In its simplest form, the MIP model of the facilities location problem can be stated as (similar formulations appear in [1, 6, 9, 10, 11, 24, 26, 27, 31, 36, 39, 40]):

Received September 1975; revised September 1976.

$$P: \quad \text{minimize} \sum_{k \in K} \sum_{j \in J} c_{kj} x_{kj} + \sum_{k \in K} f_k y_k \qquad (1)$$

$$\text{subject to} \sum_{k \in K} x_{kj} = D_j \qquad \forall j \in J \qquad (2)$$

$$\sum_{j \in J} x_{kj} \leqslant Q_k y_k \qquad \forall k \in K \qquad (3)$$

$$x_{kj} \geqslant 0 \qquad \forall k, j \qquad (4)$$

$$y_k = 0, 1 \qquad \forall k \qquad (5)$$

where K = set of indices of available sites, $K = \{1, 2, \ldots, m\}$

J = set of indices of demand centers or destinations, $J = \{1, 2, \ldots, n\}$

Q_k = capacity of a facility at $k \in K$

D_j = demand at destination $j \in J$

f_k = cost of opening a facility at $k \in K$

c_{kj} = cost for producing one unit at site $k \in K$ and delivering it to destination $j \in J$

x_{kj} = quantity produced at site $k \in K$ and delivered to destination $j \in J$

$$y_k = \begin{cases} 1 \text{ if a facility is opened at site } k \in K, \\ 0 \text{ otherwise.} \end{cases}$$

The constraints (2) require that each demand be satisfied, while (3) restricts the output of facility k to no greater than Q_k. Q_k could be made arbitrarily large if there are no real capacity restrictions.

Clearly, P represents a highly simplified model of the facilities location problem. Geoffrion [18] has described several aspects of the general problem which are not incorporated in P. In the general problem, there may be several stages of production and distribution. There may be economies of scale or other nonlinearities in the variable production and distribution costs. It may be required that a customer is served from only one facility. If storage-in-transit is a factor, then the units delivered to a customer should be identified by their point of origin. Finally, there may be other constraints on system configuration, such as a limit on the number of sites chosen from some subset.

Even so, problem P has been a popular subject for study, primarily because it has a structure to which general purpose algorithms may readily be adapted. One aspect of this special structure becomes apparent on rearranging P:

$$P: \underset{\text{s.t. (5)}}{\text{Minimize}} \left\{ \sum_{k \in K} f_k y_k + \underset{\substack{\text{s.t.(2),} \\ \text{(3), and (4)}}}{\text{minimum}} \sum_{k \in K} \sum_{j \in J} c_{kj} x_{kj} \right\} \qquad (6)$$

The inner minimization, or subproblem, can be solved as a transportation problem if the constraints (3) are binding, or by solving n one-row optimizations (each assigns a demand to the least cost source) otherwise. In the latter case, an efficient solution procedure was described by Efroymson and Ray [9]. In the former case, any of several very powerful solution procedures may be employed [4, 23]. Let $T[A]$ denote the optimal value for the subproblem with $y_k = 1$, $k \in A$, $y_k = 0$, $k \notin A$, where A is any subset of K (if A generates an infeasible solution to P, $T[A] = \infty$).

This partitioning of problem P is the basis for solving P using implicit enumeration algorithms [2, 32]. That is, P may be solved by considering (explicitly or implicitly) all the 0-1 vectors, y, and for each one, solving the subproblem in order to evaluate the vector.

Another aspect of the special structure of P may be seen by relaxing the integer restrictions on the y_k, i.e., relaxing the constraints (5) to

$$0 \leqslant y_k \leqslant 1 . \qquad (7)$$

Efroymson and Ray [9] observed that in this case the constraints (3) would hold as equalities at optimality. Thus y_k could be eliminated and the relaxed problem could be solved as:

$$\bar{P}: \quad \text{minimize} \sum_{k \in K} \sum_{j \in J} [c_{kj} + f_k/Q_k] x_{kj} \qquad (8)$$

$$\text{subject to} \sum_{j \in J} x_{kj} \leqslant Q_k \quad \forall k \in K \qquad (9)$$

(2) and (4).

Note that $\bar{P}$ has the same form as the subproblem in (6), i.e., it is a transportation problem. The relaxed problem, $\bar{P}$, is the basis for solving P by the branch and bound method [5, 29].

The special structure of P permits both the subproblems in implicit enumeration and the relaxed problem in branch and bound to be solved as transportation problems. Thus, P, can be solved using procedures more efficient than a general LP algorithm for solving the subproblems in the optimization. This means that specialized enumeration algorithms for solving problem P will be potentially much faster than general purpose algorithms.

Enumeration Procedures for Problem P

Some terminology and notation are needed to simplify the following discussion. *CP* will denote a candidate problem, that is, the problem P or a restricted version of P. The restrictions are in the form of constraints on some of the 0-1 variables, y_k, which fix their values. Let K_0 be the subset of $k \in K$ such that $y_k = 0$ is required for the *CP*, and let K_1 be the subset of $k \in K$ such that $y_k = 1$ is required. Then K_2 is the subset of $k \in K$ for which y_k is free to take on either value.

The candidate list is simply a list of all the candidate problems which have not yet been eliminated or fathomed. A given *CP* is fathomed by one of several results: it is solved;

it is shown to have no feasible solution; or it is shown to have no feasible solution with a smaller value than $\bar{Z}$, the value of the current incumbent (i.e., the best solution found so far). The latter condition is satisfied if a lower bound on $v(CP)$, e.g., $v(\overline{CP})$, can be found which is larger than $\bar{Z}$. If a CP cannot be fathomed, then some variable y_k, $k \in K_2$ is chosen to be restricted, and the CP is separated into two new CP's, corresponding to the two restrictions $y_k = 0$ and $y_k = 1$. The basic outline of the general procedure is illustrated in Fig. 1 (for a general discussion see [21]).

The following discussion of algorithmic devices and strategies refers to the papers cited in this survey and is by no means intended to deal with mixed integer programming in general.

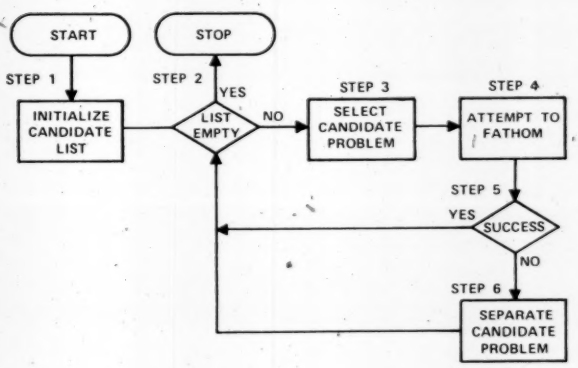


Fig. 1. General form of enumerative algorithms.

Fathoming Methods

Many problem specific devices have been developed in an attempt to enhance step 4 of the procedure, i.e., fathoming the CP. Several of these devices will now be described.

The Δ-simplification [9, 26] asks the question, "For a given CP, what is the smallest possible cost increase resulting from the restriction, $y_k = 0$ for $k \in K_2$?" If this minimum impact is nonnegative, then an optimal solution to the CP must have $y_k = 1$. For the given CP, suppose $y_k = 1, k \in K_1 \cup K_2$. The minimum increase in variable cost when $y_k = 0$ for some $k \in K_2$ occurs if all other sites in K_2 remain open. Traded off against this variable cost increase is a savings of f_k in the fixed costs incurred. Thus, the smallest possible cost increase is:

$$\Delta_k = T[K_1 \cup K_2 - \{k\}] - T[K_1 \cup K_2] - f_k \quad k \in K_2 \ . \qquad (10)$$

If $\Delta_k \geqslant 0$ then $y_k = 1$ in the optimal solution to CP.

In a similar fashion, the Ω-simplification [9, 26] asks the question, "For a given CP, what is the largest possible cost reduction resulting from the restriction $y_k = 1$ for $k \in K_2$?" If this maximum impact is nonpositive then an optimal solution to CP must have $y_k = 0$. For the given CP, suppose $y_k = 0$ for $k \in K_0 \cup K_2$. The largest possible reduction in variable cost associated with setting some $y_k = 1$ occurs if all other sites $k \in K_2$ remain closed. Traded off against this reduction in variable cost is an increase of f_k in the fixed costs incurred. Thus, the largest possible cost saving is:

$$\Omega_k = T[K_1] - T[K_1 \cup \{k\}] - f_k, \quad k \in K_2 \ . \qquad (11)$$

If $\Omega_k \leqslant 0$ then $y_k = 0$ in the optimal solution to CP. Note that the condition $\Omega_k \leqslant 0$ can only be satisfied if K_1 generates a feasible solution, i.e., $T[K_1] < \infty$.

The maximum useful capacity simplification [9, 26] asks the question, "For a given CP, are there any destinations $j \in J$ which would never be served from site $k \in K_2$ in an optimal solution to CP?" If all the corresponding demands are disregarded at site k, then the remaining demands represent the maximum useful capacity, $\hat{Q}_k$. In the event that $\hat{Q}_k < Q_k$, the lower bound for the candidate problem given by $v(\overline{CP})$ can be refined by replacing Q_k by $\hat{Q}_k$ in (8). If any $x_{kj} > 0$ for $k \in K_2$, then $v(\overline{CP})$ will be at least as large, and may be larger, with correspondingly greater chances of fathoming by having $v(\overline{CP}) > \bar{Z}$. One way to determine $\hat{Q}_k$ is to solve for $T[K_1]$, and compare c_{kj} to the (largest) cost for satisfying D_j, for all $j \in J$ and $k \in K_2$. If c_{kj} is larger than the cost used in $T[K_1]$, then site k will not be a candidate to supply destination j in the optimal solution to CP.

These three simplifications were first used by Efroymson and Ray [9] in an algorithm for the problem without capacity restrictions. For that problem, $T[A]$ can be determined by inspection (i.e., assign each D_j to site ℓ where $c_{\ell j} = \min_{k \in A} c_{kj}$) so that Δ_k and Ω_k could be computed as in (10) and (11). For the case of binding capacity restrictions, applying (10) and (11) to a particular CP would require solving $2|K_2|+2$ transportation problems ($|K_2|+1$ for each). Even though there are quite good algorithms for the transportation problem, this could prove too costly in terms of computing effort. For this reason, Akinc [1] and Khumawala and Akinc [27] developed approximations to Δ_k and Ω_k. The Δ_k approximation was derived by relaxing all capacity restrictions in the subproblems, except the one associated with site k, which again permits the Δ_k to be determined by inspection. The Ω_k approximation was derived by partially dualizing the demand constraints (similar to Lagrangian relaxation, see [17]) which allows it to be determined by solving a continuous knapsack problem.

For a given CP, a lower bound on variable cost in any solution is $T[K_1 \cup K_2]$. Thus, given the optimal y_k, the value of any solution satisfies

$$v(CP) \geqslant \sum_{k \in K} f_k y_k + T[K_1 \cup K_2] \ . \qquad (12)$$

If it can be shown that $v(CP) \geqslant \bar{Z}$, then the CP may be fathomed. This inequality combined with (12) yields a constraint on the optimal y_k:

$$\sum_{k \in K} f_k y_k \leqslant \bar{Z} - T[K_1 \cup K_2] \ . \qquad (13)$$

If (13) will be violated by adding the restriction $y_k = 1$, $k \in K_2$, then clearly $y_k = 0$ is required for the optimal solution of *CP* to improve the incumbent solution. A weaker form of (13) results by noting that $T[K] \leqslant T[K_1 \cup K_2]$ and replacing $T[K_1 \cup K_2]$ by $T[K]$.

Inequality (13) is a fixed cost bound, *FCB*, and was originally suggested by Gray [24]. It has subsequently been used in most algorithms, since it is easy to implement, especially in the weaker form.

A bound similar in spirit to (13) can be derived from the Ω-simplification. If K_1 generates a feasible solution, let

$$Z_{CP} = T[K_1] + \sum_{k \in K_1} f_k \qquad (14)$$

denote the value of the solution to *CP* when all $y_k = 0$, $k \in K_2$. Recall that Ω_k is the maximum reduction in Z_{CP} for setting $y_k = 1$. Therefore, the largest possible reduction in Z_{CP} can be found by summing the positive Ω_k. In order for *CP* to lead to an improved solution, the lower bound must not exceed $\bar{Z}$, i.e.,

$$Z_{CP} - \sum_{k \in K_2} \max(0, \Omega_k) \leqslant \bar{Z}. \qquad (15)$$

If (15) is violated, then the optimal solution of *CP* cannot improve the incumbent, so *CP* may be fathomed. A version of this Ω-bound was first proposed by Spielberg [39].

Spielberg also discussed the use of another type of inequality, originally developed by Benders [3]. If $u = (u_i)$ is a feasible solution vector for the dual of the subproblem for some set of fixed y_k, and $\bar{Z}$ is the value of the incumbent, then the optimal solution to *P* must satisfy the Benders inequality:

$$\sum_{k \in K} [f_k - Q_k(u_k - u_s)]\, y_k \leqslant \bar{Z} - \sum_{j \in J} D_j(u_j - u_s). \qquad (16)$$

(The index s denotes a dummy destination used to absorb excess capacity.) Suppose that the fixed y_k values lead to an improved feasible solution. Then the inequality holds as an equality for *that set* of y_k, and the first summation can be thought of as the cost of capacity while the second can be thought of as the cost of meeting demands. An inequality (16) may be generated each time the subproblem is solved with a different set of fixed y-values.

Each constraint (16) limits the sets of feasible values which need to be considered in a search for the optimum solution. Individually, however, the constraints are weak in the sense that they are satisfied by many feasible solutions. Spielberg demonstrated that the Ω-bound is as strong an inequality for the solutions of a given *CP* (assuming it is capacity-feasible) and so did not actually employ (16) in his algorithm. The Benders inequalities are explicitly used in the algorithm of Ellwein and Gray [10].

In an attempt to overcome the inherent weakness of the individual Benders inequalities, Bulfin [6], and Bulfin and Unger [7] employed a surrogate constraint (see [2, 16, 22] for a discussion of surrogate constraints). The essence of the surrogate constraint technique is to form a linear combination of inequalities, and thereby capture the logical requirements of several constraints in a single inequality. Bulfin determined a "best" surrogate using the procedure developed by Geoffrion [16], which requires solving a particular LP problem to generate the surrogating coefficients.

Rather than seek special bounds or dominance conditions, Davis and Ray [8] reformulated the problem. They replaced the constraints (3) by (9) and replaced (4) by:

$$0 \leqslant x_{kj} \leqslant y_k \min(D_j, Q_k) \quad \forall k \in K, \ \forall j \in J. \qquad (17)$$

The relaxed problem for this reformulation could not be solved as a transportation problem because the y_k could not be eliminated as in $\bar{P}$. Thus, a linear programming procedure was required. As a reward for the added computational burden, their relaxed problem was found to terminate integer in most cases, so that for a given problem, very few candidate problems (sometimes only one) had to be examined.

Rardin [33] and Rardin and Unger [34] used a group theoretic approach to develop strong penalty bounds for use in a branch and bound algorithm for solving problem *P*. After solving for $v(\overline{CP})$, they attempted to improve the lower bound by forming the corresponding group problem, which they then relaxed to obtain penalty bounds.

The devices described so far have been aimed at improving the execution of step 4 in the general procedure, i.e., attempting to fathom the *CP*. The effect of alternate strategies in steps 3 and 6, selecting the next *CP* and separating the current *CP*, have also been examined.

Selecting the Candidate Problem

Several rules have been used for selecting the next *CP*, which will be referred to as node selection rules. The traditional rule for branch and bound algorithms is "least lower bound," or LLB. The traditional rule for implicit enumeration algorithms is "last-in-first-out," or LIFO. The use of a LIFO rule simplifies the implementation of the logical tests usually found in implicit enumeration algorithms. These are both special cases of a general "priority" rule. Other possible priorities are discussed in [5, 14, 21, 32].

LLB and LIFO are two of the most often used rules. Each rule has its advantages and disadvantages. A well-known difficulty with LLB is that it may exhaust the available space for the candidate list. LIFO does not present such a space problem, but is much less flexible with regard to which *CP* may be considered next. Akinc [1] developed a hybrid rule to overcome the space limitations of LLB without giving up too much of its flexibility. His rule is: "Use LLB until the size of the candidate list exceeds β, then use LIFO until the size of the candidate list falls to α," where α and β are parameters determined *a priori* by considering the space available for the list. Similarly, the rule used by Ellwein and Gray [10] attempts to overcome

the inflexibility of LIFO by allowing the next *CP* to be *any* *CP* which differs from the current *CP* in exactly one element of K. This NSR will be called the "single swap" rule.

Selecting the Branching Variable

Various rules for choosing the separation, or *branching*, variable have been formulated and tested. These *branch selection rules*, BSR, can be stated as: "select the free variable with the largest (smallest) evaluation."

Efroymson and Ray [9] suggested using as an evaluator the maximum useful capacity, $\hat{Q}_k$ (they described the rule in terms of the maximum demand which could be served from each site). This rule is motivated by a desire to make the *CP* capacity feasible as quickly as possible. Sá [36] used the values of $\bar{y}_k$, the fractional solution to $\overline{CP}$, as evaluators, while Davis and Ray used $|y_k - .5|$. Spielberg suggested a rule which essentially used Δ_k as the evaluator.

Khumawala [26] and Akinc [1] gave consideration to four different rules, and compare their performances on a set of problems. They used as evaluators, Δ_k, Ω_k, $\hat{Q}_k$, and $\bar{y}_k$, and considered both maximum and minimum evaluators.

When the Benders inequalities are used, a candidate for the evaluator is largest or smallest change in infeasibility for the current set of inequalities. This rule is used by Ellwein and Gray [10] and Bulfin [6].

Computational Results for Problem *P*

It is of interest to determine the extent to which the previously described devices may be successfully used in solving facilities location problems. This is a difficult task, due to the methods used to evaluate procedures and the reporting of companion results.

The empirical results discussed below all relate to a set of "standard" problems, which are based on the data given by Kuehn and Hamburger [28]. Results for the seven algorithms described in Table 1 will be given. It should be noted that the algorithms of Ellwein and Gray [10], McGinnis [31], and Soland [38] are designed for more complex models than *P*, so their performance might be improved by specializing certain steps. The branching rule in Soland's algorithm selects the site that maximizes

$$f_k - \frac{f_k}{Q_k} \sum_j x_{kj} \tag{18}$$

which is the difference between the fixed cost and the amount of the fixed cost incurred in the solution of the relaxed problem.

Table 2 summarizes the CPU times reported by the various authors for several versions of the test problems. Note that the capacities and fixed costs are the same at all sites. The row labeled (Open Init./Open Opt.) refers to the number of sites which could be fixed open by applying the Δ- and Ω-simplifications before beginning the enumeration. The 16 X 50 and 25 X 50 problems have, respectively, 16 sites and 50 demand centers, and 25 sites and 50 demand centers. For the usual reasons (i.e., differences in machines and coding optimization), it is not meaningful to make direct comparisons of CPU times for different algorithms. However, a careful examination of Table 2 leads to some interesting conclusions regarding both types of algorithms and data dependencies.

Without making direct comparisons of solution times, there appears to be a dominance of branch and bound algorithms over the implicit enumeration procedures, simply because more problems were solved by branch and bound. Of course, this might simply reflect a relatively smaller empirical effort for the latter. A similar general observation is that the uncapacitated (Q_k = 58268 = total demand $\forall\, k$) problems are "well-solved," compared to the problems with binding capacity restrictions.

For problems with given site capacities, say 15000 units, the level of fixed costs has a definite effect on solution times. In general solution times for a given algorithm increase with increasing fixed costs up to a point, and then decrease. This indicates that if fixed costs are very small or very large the problem will be easier to solve.

This observation is reinforced by considering the Δ and Ω simplification. If f_k is small enough so that $\Delta_k \geq 0$ or large enough so that $\Omega_k \leq 0$, then y_k can be fixed. Thus,

Table 1: Algorithms reported in Table 2.

Author		Description of Algorithm										Computer
		$\overline{CP}$	Δ_k	Ω_k	Ω_k	*FCB*	Ω-*B*	*BI*	*SC*	*NSR*	*BSR*	
Akinc	[1]	Y	Y	Y	Y	N	N	N	N	Hybrid	max Ω_k	CDC 6500
Bulfin	[6,7]	N	Y(a)	N	N	Y	N	Y	Y	LIFO	min inf	U 1108
Ellwein-Gray	[10]	N	Y(a)	N	N	Y	N	Y	N	Single Swap	min inf	IBM 360/65
Khumawala	[26]	Y	Y	Y	Y	N	N	N	N	LLB	max Ω_k	CDC 6500
McGinnis	[31]	N	N	N	N	Y	N	Y	Y	LIFO	min inf	IBM 370/165
Sá	[36]	Y	N	Y	N	Y	N	N	N	LIFO	min (max) y_k	IBM 360/65
Soland	[38]	Y	N	N	N	N	N	N	N	LLB	max error	CDC 6600

(a) – initial node only

Table 2: Computational results for standard problems (CPU times in seconds).

Capacities Q_k		5000				15000				58268			
Fixed Costs f_k		7.5	12.5	17.5	25	7.5	12.5	17.5	25	7.5	12.5	17.5	25
(Open Init./Open Opt.)		(11/12)	(11/12)	(11/12)	(11/12)	(9/11)	(7/9)	(3/7)	(3/5)	(9/11)	(7/9)	(3/5)	(3/4)
Sá	[36]	90.6	108.0	96.0	87.6	94.2	457.2	900+	–	–	–	–	–
Soland	[38]	38.7	5.8	4.2	–	–	271.3	–	–	–	–	–	–
Akinc	[1]	10.21	9.15	9.26	9.58	0.23	0.43	38.65	34.39	47.52	0.44	0.30	0.15
Ellwein & Gray	[10]	–	88.8	92.4	–	–	127.8	–	–	–	15.0	63.0	–
Bulfin	[6]	–	27.0	25.0	–	–	87.0	–	–	–	43.8	28.2	–
McGinnis	[31]	2.1	2.5	2.8	3.5	–	–	–	–	–	–	–	–
16 X 50 problems													
(Open Init./Open Opt.)		(11/17)	–	–	–	(10/15)	(6/11)	–	–	(10/15)	(6/11)	(4/8)	(2/4)
Khumawala	[26]	–	–	–	–	–	–	–	–	1.50	1.36	1.66	0.85
Akinc	[1]	0.23	120+	120+	120+	0.75	7.75	120+	120+	0.68	1.65	1.34	0.46
Ellwein & Gray	[10]	–	–	–	–	–	–	–	–	196.2	–	–	–
Bulfin	[6]	–	–	–	–	–	–	–	–	123.0	–	–	–
25 X 50 problems													

+ : time limit exceeded

there is an intermediate range of values for f_k which do not allow the simplifications to occur:

$$T[K_1] - T[K_1 \cup \{k\}] > f_k > T[K_1 \cup K_2 - \{k\}] - T[K_1 \cup K_2]. \tag{19}$$

In other words, if

$$\begin{bmatrix}\text{maximum variable cost} \\ \text{reduction for site } k \text{ open}\end{bmatrix} > f_k > \begin{bmatrix}\text{minimum variable cost} \\ \text{reduction for site } k \text{ open}\end{bmatrix} \tag{20}$$

for most $k \in K_2$, the problem should be relatively harder to solve.

The level of site capacities also affects ease of solution. Recall that the Ω-simplification can't be used until the *CP* is capacity feasible, i.e., total capacity of open sites exceeds total demands. If the Q_k are small relative to total demands, then there may be many *CP* in the candidate list when the Ω simplification begins to have an effect. It is tempting to argue that problems will be hard to solve when about one-half the sites must be open in a feasible solution. However, Table 2 does not seem to support such an argument.

In summary, it appears that the uncapacitated problems are easily solved, but the large problems with capacity constraints are not, unless fixed costs are small.

Other researchers have reported computational results for problems other than the Kuehn-Hamburger set. Spielberg was able to solve randomly generated uncapacitated problems with 20 to 100 sites in 1 to 75 minutes on an IBM 360/40 [39] and 52-100 sites in 2 to 167 minutes on an IBM 360/50 [40]. Rardin [33] solved some randomly generated problems on a Univac 1108 with results as summarized in Table 3.

Table 3: Rardin's computational results.

Sources	Destinations	% Infeasible Routes ($c_{kj} = \infty$)	Solution Time (sec.)
50	10	50	4.2 - 242
75	15	67	19.4 - 478.4
100	20	75	22.2 - 803.6

Related Results

Two closely related generalizations of problem *P*, which have received attention are the case of economies of scale (or other nonlinearities) in the variable costs and the case of "configuration constraints."

There are two basic approaches to dealing with nonlinear facility costs. One is to *a priori* form a piecewise linear estimate of the cost function, then let there be a number of "pseudofacilities" corresponding to the number of linear segments. It will usually be necessary to restrict the throughput of each pseudofacility to the range associated with the corresponding linear cost segment, and to add a constraint to prevent more than one pseudofacility to be open at any site. This is the method used by Ellwein and Gray [10], Elshafei [11], and McGinnis [31] in their implicit enumeration algorithms.

The second approach is very similar to the first, in that it uses piecewise linear underestimates of the original cost function. However, rather than define the approximation *a priori*, the underestimates are refined as required to satisfy a preset tolerance for approximation error. This method has been implemented by Soland [38], Rech and Barton [35], and Florian, Robillard and Achim [13] in the framework of branch and bound.

There are very few computational results reported for these procedures in solving problems with scale economies. In the first group, the largest problems reported have 17 sites and 39 pseudofacilities, and required less than one minute on an IBM 370/165 [31]. In the second group, the largest facilities location problems (with scale economies or other nonlinearities) solved had 10 sites [35]. Florian, Robillard and Achim [13] report solving some concave cost network flow problems with as many as 55 concave cost arcs. (A facilities location problem can be cast as a concave cost flow problem by letting an arc rather than a node, represent the site.)

Problem P with additional (configuration) constraints and lower as well as upper bounds on site capacities has been solved by Geoffrion and McBride [20]. They implemented a penalty bound based on Lagrangian relaxation of the CP [12, 17] and reported that the bound obtained was considerably stronger than $v(\overline{CP})$, while requiring only a modest amount of additional computation. Problems with 7 to 24 sites were solved in less than 20 seconds on an IBM 360/91. It should be noted that the procedures using an *a priori* piecewise linear estimate of a nonlinear cost function incorporate some special cases of general configuration constraints.

Geoffrion and Graves [19] considered an even more general problem which had two stages of delivery, no customer serviced by more than one supply point, and deliveries to customer identified by point of origin. The solution procedure they describe is a variant of Benders partitioning algorithm, and was capable of solving problems with up to 30 sites in 17 to 191 seconds on an IBM 360/91.

Promising Developments

There appear to be at least three approaches to improving the algorithms available for solving problem P. One approach is to improve the efficiency of the imbedded optimizations. Examples of this would be the incorporation of better network codes [23] or better LP codes [37]. This approach has only limited potential because it does not affect the number of imbedded optimizations.

A second approach is the application of general developments, such as Lagrangian relaxation [17]. This has already proven to be a fertile area for improvement. The use of group-based penalties [33] and more sophisticated rules for candidate problem selection and separation [5, 14, 32] may yield further improvement.

A third approach is the continued development of problem-specific devices, such as the Δ- and Ω-simplification. As the state-of-the-art in optimization advances, more sophisticated dominance tests become worthwhile, provided they are effective.

Conclusion

Much of the previous work with MIP procedures for problem P has concentrated on developing problem specific devices to improve the fathoming process. This effort has not been consistently successful in permitting the solution of large (25 sites) problems with binding capacity restrictions. Several new developments of a general nature give promise of overcoming some of the current limitations, although extensive testing has yet to be reported.

It is more difficult to assess the situation when consideration is extended to problems exhibiting economies of scale in the variable production costs or problems subject to "configuration constraints." The main reason for this is that although algorithms have been proposed for these problems, there are few computational results available. At best, the capability for solving these problems, in terms of the number of sites considered, is no more advanced than for problem P.

The facilities location problem has received considerable attention in the past, but is still not a "well-solved" problem. Further research seems warranted.

Acknowledgment

The author wishes to thank Dr. S. E. Elmaghraby, who introduced him to the problems of facilities location.

References

[1] Akink, U., "A Branch and Bound Procedure for Solving Warehouse Location Problems with Capacity Constraints," Unpublished PhD dissertation, University of North Carolina, Chapel Hill (1973).

[2] Balas, E., "Discrete Programming by the Filter Method," *Operations Research*, **15**, 915-957 (1967).

[3] Benders, J. F., "Partitioning Procedures for Solving Mixed Variables Programming Problems," *Numerische Mathematik* **4**, 238-252 (1962).

[4] Bradley, G. H., G. G. Brown, and G. W. Graves, "A Comparison of Storage Structures for Primal Network Codes," Spring 1975 ORSA/TIMS National Meeting.

[5] Breu, R., and C. Burdet, "Branch and Bound Experiments in Zero-One Programming," *Mathematical Programming Study*, **2** 1-50 (1974).

[6] Bulfin, R. L., "An Algorithm for the Plant Location Problem," Unpublished Master's Thesis, Georgia Institute of Technology, Atlanta, Georgia (1972).

[7] Bulfin, R. L., and V. E. Unger, "Computational Experience with an Algorithm for the Lock Box Problem," *Proc. ACM*. 16-19 (1973).

[8] Davis, P. S., and T. L. Ray, "A Branch-Bound Algorithm for the Capacitated Facilities Location Problem," *Naval Research Logistics Quarterly* **16**, 331-343 (1969).

[9] Efroymson, M. A., and T. L. Ray, "A Branch and Bound Algorithm for Plant Location," *Operations Research*, **14**, 361-368 (1966).

[10] Ellwein, L. B., and P. L. Gray, "Solving Fixed Charge Location-Allocation Problems with Capacity and Configuration Constraints," *AIIE Transactions*, **3**, 290-299 (1971).

[11] Elshafei, A. N., "Optimal Location of Facilities," Unpublished PhD dissertation, University of Birmingham, Birmingham, England (1973).

[12] Fisher, M. L., W. D. Northrup, and J. F. Shapiro, "Using Duality to Solve Discrete Optimization Problems: Theory and Computational Experience," *Mathematical Programming Study*, **3**, 59-94 (1975).

[13] Florian, M.,P. Robillard, and C. Achim, "Solving Large-Scale Concave Cost Network Flow Problem," Publication #103, Department d'informatique, Université de Montréal (1972).

[14] Forrest, J. J. H., J. P. H. Hirst, and J. A. Tomlin, "Practical Solution of Large Mixed Integer Programming Problems with UMPIRE," *Management Science*, **20**, 736-773 (1974).

[15] Garfinkel, R. S., and G. S. Nemhauser, *Integer Programming*, John Wiley & Sons, New York (1972).

[16] Geoffrion, A. M., "An Improved Implicit Enumeration Approach for Integer Programming," *Operations Research*, **17**, 437-454 (1969).

[17] Geoffrion, A. M., "Lagrangian Relaxation and Its Uses in Integer Programming," *Mathematical Programming Study*, **2**, 82-114 (1974).

[18] Geoffrion, A. M., "A Guide to Computer-Assisted Methods for Distrubtion Systems Planning," *Sloan Management Review*, **16**, 17-42 (1975).

[19] Geoffrion, A. M., and G. Graves, "Multicommodity Distribution System Design by Benders' Decomposition," *Management Science*, **20**, 822-844 (1974).

[20] Geoffrion, A. M., and R. D. McBride, "The Capacitated Facility Location Problem with Additional Constraints," AIIE, ORSA/TIMS Joint National Meeting (1973).

[21] Geoffrion, A. M., and R. E. Marsten, "Integer Programming Algorithms: A Framework and State-of-the-Art Survey," *Management Science*, **18**, 465-491 (1972).

[22] Glover, F., "Surrogate Constraints," *Operations Research*, **16**, 741-749 (1968).

[23] Glover, F., D. Karney, D. Klingman, and A. Napier, "A Computational Study on Start Procedures, Basis Change Criteria, and Solution Algorithms for Transportation Problems," *Management Science* **20**, 793-813 (1974).

[24] Gray, P., "Mixed Integer Programming Algorithms for Site Location and Other Fixed Charge Problems Having Capacity Constraints," Technical Report No. 101, Stanford University, California (1967)

[25] Hakimi, S. L., and S. N. Maheshwari, "Optimum Locations of Centers in Networks," *Operations Research*, **20**, 967-973 (1972).

[26] Khumawala, B. M., "An Efficient Branch and Bound Algorith for the Warehouse Location Problem," *Management Science*, **18**, 718-731 (1972).

[27] Khumawala, B. M., and U. Akinc, "An Efficient Branch and Bound Algorithm for the Capacitated Warehouse Location Problem," Paper No. 475, Krannert Graduate School of Industrial Administration, Purdue University, Lafayette, Indiana (1974); forthcoming in *Management Science*.

[28] Kuehn, A. A., and M. J. Hamburger, "A Heuristic Program for Locating Warehouse," *Management Science*, **9**, 643-666 (1963).

[29] Lawler, E. L., and D. E. Wood, "Branch and Bound Methods: A Survey," *Operations Research*, **14**, 699-719 (1966).

[30] Lea, A. C., "Location-Allocation Systems: An Annotated Bibliography," Discussion Paper No. 13, Department of Geography, University of Toronto, Canada (1973).

[31] McGinnis, L. F., "Approximate and Exact Solution Procedures for a Class of Facilities Location Problems, PhD dissertation, North Carolina State University, Raleigh, North Carolina (1975).

[32] Piper, C. J., "Implicit Enumeration: A Computational Study," Working Paper Series No. 115, School of Business Administration, University of Western Ontario, London, Canada (1974).

[33] Rardin, R. L., "Group Theoretic and Related Approaches to Fixed Charge Problems," PhD dissertation, School of Industrial and Systems Engineering, Georgia Institute of Technology, (1973).

[34] Rardin, R. L., and V. E. Unger, "Solving Fixed Charge Network Problems with Group Theory-Based Penalties," *Naval Research Logistics Quarterly*, **23**, 67-84 (1976).

[35] Rech, P., and L. G. Barton, "A Non-Convex Transportation Algorithm," in E. M. L. Beale (ed.), *Applications of Mathematical Programming Techniques*, American Elsevier Publishing Company, pp. 250-260 (1970).

[36] Sa, G., "Branch-and-Bound Approximate Solutions to the Capacitated Plant-Location Problem," *Operations Research*, **17**, 1005-1016 (1969).

[37] Schrage, L., "Implicit Representation of Variable Upper Bounds in Linear Programming," *Western Management Science Institute*, UCLA (1974).

[38] Soland, R. M., "Optimal Plant Location with Concave Costs," *Operations Research*, **22**, 273-382 (1974).

[39] Spielberg, K., "Plant Location with Generalized Search Origin," *Management Science*, **17**, 165-177 (1969).

[40] Spielberg, K., "Algorithms for the Simple Plant-Location Problem with Some Side Conditions," *Operations Research*, **17**, 85-111 (1969).

[41] Wendell, R. E., and A. P. Hurter, "Optimal Location on a Network," *Transportation Science*, **7**, 18-33 (1973).

Dr. Leon F. McGinnis, Jr., is Assistant Professor of Industrial and Systems Engineering at Georgia Institute of Technology. His research activities include applications of operations research methodologies to problems in facility location, scheduling and inventory systems. He received the BIE from Auburn University and the MSIE and PhD degrees from N. C. State. Dr. McGinnis is a member of AIIE, ASEE, ORSA and TIMS

A Comparison of Two Optimal Inspection Procedures in the Presence of Errors

MATTHEW ROSENSHINE
SENIOR MEMBER, AIIE
The Pennsylvania State University
University Park, Pennsylvania 16802

FRANCISCO SOARES
General Electric Co.
Sao Paulo, Brazil

Abstract: It is well known that the Bayes procedure minimizes the expected loss when the a priori probabilities of the states of nature are known, but it is not clear that the procedure continues to have this property in the presence of unknown errors which occur if these probabilities must be estimated. Furthermore, it is possible that the expected loss for another procedure does not increase as fast as that for the Bayes procedure. A procedure which might behave this way is the procedure for a group of N items, designated the N-procedure, because the expected loss for this procedure does not depend explicitly on the a priori probabilities. It is found, however, that the Bayes procedure consistently yields lower expected costs than the N-procedure. The results obtained suggest that the Bayes procedure will give smaller expected losses than other procedures in the presence of unknown errors in the a priori probabilities. Somewhat surprisingly, the superiority of the Bayes procedure increases with the size of the errors of the estimated a priori probabilities, regardless of whether the estimated probabilities lie above or below the correct values.

■ There are several binary (two-state) decision procedures, that can be used to decide which of two states-of-nature exist. It is known that the Bayes procedure minimizes the expected loss when the a priori probabilities of the states of nature are known. However, when the a priori probabilities must be estimated, little is known about the relative behavior of the several procedures. This paper represents an attempt to shed some light on the behavior of binary procedures by examining the relative behavior of two binary procedures in the presence of errors in the a priori probabilities.

The procedures chosen for comparison are two of the five procedures presented in [1] – the Bayes procedure and the procedure for a group of N-items (designated the N-procedure). The reasons for this choice are simple. If expected loss is chosen as the measure of a procedure's effectiveness, then the Bayes procedures is a likely candidate to continue its relative superiority. On the other hand, the only one of the five procedures for which the expected loss does not depend explicitly on the a priori probabilities is the N-procedure. Thus, it is not unreasonable to hypothesize that the effectiveness of this procedure will be affected less by errors in the a priori probabilities than the Bayes procedure.

The practical implications of the comparison presented in this paper are best illustrated by an example. In many instances, inspection plans which are designed to differentiate between two states of nature (acceptable and unacceptable with respect to some attribute) are not based upon the measurement of the attribute sought, simply because this attribute is not directly measurable. In such cases, another attribute, which can be measured, and which is correlated with the attribute sought, becomes the basis for the inspection plan. For example, the acceptability of steel bars with regard to tensile strength (attribute sought) is frequently decided by the measurement of the hardness (measurable attribute) of the bars. The numerical values of the measurable attribute, which satisfy the sought attribute requirements, can be regarded as the values assumed by a random variable, because there exists a correlation rather than a deterministic relationship between it and the

Received September 1975; revised February 1976; July 1976.

attribute sought. By the same reasoning, it can also be said that the numerical values of the measurable attribute, which do not meet the sought requirements, are the values assumed by another random variable. Thus, it is seen that binary decision procedures are often called for in industrial applications. In such cases, the practitioner needs to know what decision procedure he should use. If the a priori probabilities are not known without error, the answer is not clear.

Description of Procedures

The Bayes procedure minimizes the expected loss, $\bar{L}_B$, which depends on the a priori probability of each state-of-nature and the losses associated with the various decision, state-of-nature combinations. The minimization of $\bar{L}_B$ is accomplished by accepting those items with measurements, x, satisfying

$$\frac{f(x \mid G)}{f(x \mid U)} > \frac{(1-P_G)\ (L_{GU}-L_{UU})}{P_G(L_{UG}-L_{GG})}$$

where

P_G = a priori probability that an item is good

p.d.f. = probability density function

$f(x|G)$ = p.d.f. for population of good items

$f(x|U)$ = p.d.f. for population of unacceptable items.

L_{GG} = loss for deciding an item is good when it is good

L_{GU} = loss for deciding an item is good when it is unacceptable

L_{UG} = loss for deciding an item is unacceptable when it is good

L_{UU} = loss for deciding an item is unacceptable when it is unacceptable

Those items with measurements not satisfying the above inequality are rejected.

The N-procedure, instead of using a specified numerical threshold for the acceptance or rejection of each item, is based on rejecting some fixed number, N'_U, of items from each sample of N items. This procedure minimizes the expected loss, $\bar{L}_N$, when N'_U items must be rejected and, incidentally, maximizes the probability of rejecting all items that are actually unacceptable. It is shown in [1] that $\bar{L}_N$ is minimized if the ratios

$$\frac{f(x_j \mid G)}{f(x_j \mid U)} \qquad j = 1, 2, \ldots, N$$

are ordered and the items having the N'_U lowest ratios are rejected.

Whenever the a priori probabilities are known, the expected loss of the Bayes procedure is lower than that for the N-procedure. This occurs because the Bayes procedure considers each item out of a sample of size N "on its own merits" rather than in comparison to all the other items in the sample as the N-procedure does. Thus, in a bad (good) sample the Bayes procedure is free to reject more (fewer) items than in an "average" sample, while the N-procedure must reject N'_U items in all cases. However, if the a priori probability, P_G, is not known, an error is made in its estimation. This error results in an error in the inequality used to decide whether an item is acceptable or not and the procedure no longer guarantees that the expected loss is being minimized. On the other hand, since P_G does not enter into the N-procedure (and only enters implicitly into $\bar{L}_N$), there is a chance that $\bar{L}_N$ will not increase as fast as $\bar{L}_B$ as the error in estimating P_G increases.

The objective of this paper is to investigate whether the N-procedure can, in the presence of errors, offer a comparable or lower expected loss than the Bayes procedure. Because the expression for the expected loss for the N-procedure is very difficult to evaluate analytically, a simulation of the inspection plan was used to estimate this expected loss. The expected loss associated with the Bayes procedure is calculated analytically and can also be obtained using the simulation program (with only minor changes to reflect the Bayes procedure) to provide a check on the validity of the simulation.

The comparison of the two procedures is made assuming that the samples are drawn from exponential populations. While exponential populations may not arise in practical situations, the paper is concerned with comparing the effectiveness of the decision procedures, not with the nature of the populations on which the procedure is used. The latter should be irrelevant if the comparisons are made properly since only the degree to which one procedure is better than the other should change as the population distributions of good and unacceptable items change. Hence, exponential populations may be used without loss of generality. Furthermore, with exponential populations $[f(x|G)/f(x|U)]$ is monotonic with respect to the measurements, x, thereby simplifying many of the required calculations and the simulation computer program.

Method of Comparison

To make a fair comparison between the two procedures, they must be applied under identical conditions. Thus, N_S collections of N measurements each are chosen at random from the p.d.f.

$$P_G\ f(x|G) + (1 - P_G)\ f(x|U).$$

The Bayes procedure is applied to each of these N_S collections using EP_G as the estimate of P_G resulting in losses $L_B(1), L_B(2), \ldots, L_B(N_S)$ and mean loss $\bar{X}_B = N_S^{-1} \sum_{i=1}^{m} L_B(i)$. Then the N-procedure is applied to each of these collections for each possible value of N'_U resulting in losses

$$\left\{\begin{array}{ll} L_N & (0,1) \\ L_N & (1,1) \\ & \vdots \\ L_N & (N,1) \end{array}\right. \quad \text{for the first collection}$$

$$\left\{\begin{array}{ll} L_N & (0,2) \\ L_N & (1,2) \\ & \vdots \\ L_N & (N,2) \end{array}\right. \quad \text{for the second collection, . . . , and}$$

$$\left\{\begin{array}{ll} L_N & (0,N_S) \\ L_N & (1,N_S) \\ & \vdots \\ L_N & (N,N_S) \end{array}\right\} \quad \text{for the } N_S th \text{ collection.}$$

These loss values can be tabulated as shown in Table 1.

Table 1: Tabulation of losses.

	N'_U \ i	1	2	⋯	N_S	Sample mean	Population mean
Bayes		$L_B(1)$	$L_B(2)$		$L_B(N_S)$	$\bar{X}_B$	$\bar{L}_B$
N-procedure	0	$L_N(0,1)$	$L_N(0,2)$		$L_N(0,N_S)$	$\bar{x}_N(0)$	$\bar{L}_N(0)$
	1	$L_N(1,1)$	$L_N(1,2)$		$L_N(1,N_S)$	$\bar{x}_N(1)$	$\bar{L}_N(1)$
	2	$L_N(2,1)$	$L_N(2,2)$		$L_N(2,N_S)$	$\bar{x}_N(2)$	$\bar{L}_N(2)$
	⋮						
	N	$L_N(N,1)$	$L_N(N,2)$		$L_N(N,N_S)$	$\bar{x}_N(N)$	$\bar{L}_N(N)$

Now, if the N_S elements in each row between the double vertical lines are viewed as a sample of size N_S, then their mean, $\bar{X}_N(N'_U) = N_S^{-1} \sum_{i=1}^{N_S} L_N(N'_U, i)$, listed in the column labeled "Sample Mean," is an unbiased estimation of the quantity listed in the column labeled "Population Mean" where $\bar{L}_N(N'_U)$ is the expected loss under the N-procedure if N'_U items must be rejected. Furthermore, the standard deviation of these N_S samples can be used as a good estimate of the corresponding population standard deviation because N_S was chosen to be 500, which is quite large. Thus, the distribution of $\bar{x}_N(N'_U)$ may be assumed to be approximately normal with mean $\bar{L}_N(N'_U)$ and standard deviation $\sigma_{\bar{x}_N(N'_U)} \approx s_{L_N(N'_U, i)} / (N_S)^{1/2}$.[1]

The comparison between $\bar{L}_N(N'_U)$ and $\bar{L}_B$ is made probabilistically since $\bar{L}_N(N'_U)$ is approximated by $\bar{x}_N(N'_U)$.

$$P_r[\bar{L}_N(N'_U] < \bar{L}_B] = P_r\left[\frac{\bar{x}_N(N'_U) - \bar{L}_N(N'_U)}{\sigma_{\bar{x}_N(N'_U)}} > \frac{\bar{x}_N(N'_U) - \bar{L}_B}{\sigma_{\bar{x}_N(N'_U)}}\right]$$

$$= P_r \quad [Z > Z_{\bar{L}_B}]$$

where Z is a standard normal variable and $Z_{\bar{L}_B}$ is a known constant (because $\bar{x}_N(N'_U)$, L_B, and $\sigma_{\bar{x}_N(N'_U)}$ are all known).

For the N-procedure, the value of N'_U which should be used is the value which minimizes $\bar{L}_N(N'_U)$. This value of N'_U depends indirectly on the a priori probability P_G. If P_G is known, then using the simulation results, the best value of N'_U to use can be found from the tabulated values of $\bar{x}_N(N'_U)$ as the value of N'_U which minimizes $\bar{x}_N(N'_U)$. While this does not assure that the value of N'_U so chosen actually minimizes $\bar{L}_N(N'_U)$, there is little reason to suspect that this is not the case in view of the large value of N_S. (At worst, a value which almost minimizes $\bar{L}_N(N'_U)$ might be chosen.) This procedures results in Table 2 where appropriate values of N'_U for the indicated values of P_G (for N=100 and the parameters associated with comparison 2 of the next section) are given.

Table 2: "Best" values of N'_U.

P_G	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85
N'_U	42	40	36	32	25	14	1	0

When P_G is not known, the "best" value of N'_U is determined by using EP_G for P_G in the above table. Thus, when P_G is, say, 0.70 but is estimated at $EP_G = 0.65$, the value of N'_U that is used is 32 and the estimate of $\bar{L}_N(N'_U)$ is $\bar{x}_N(32)$ which is readily available from the simulation results; e.g., Table 4, $P_G = 0.70$ and $EP_G = 0.65$. It is this value which is compared with $\bar{L}_B$ given in [1] as

$$\bar{L}_B = P_G L_{UG} + (1 - P_G) L_{UU} + \int_{\text{Acceptance region}} [(1 - P_G)(L_{GU} - L_{UU}) f(x|U) - P_G(L_{UG} - L_{GG}) f(x|G)]\, dx$$

where $P_G = 0.70$ while using

$$\frac{f(x \mid G)}{f(x \mid U)} > \frac{(1 - EP_G)(L_{GU} - L_{UU})}{EP_G(L_{UG} - L_{GG})}$$

with $EP_G = 0.65$ to define the acceptance region.

Comparisons Made and Results

In all of the comparisons made, $f(x|G)$ and $f(x|U)$ are chosen as exponential density functions. While this choice is not necessary, it has certain advantages. It facilitates the analytical determination of $\bar{L}_B$ and provides a monotonic relationships between x and $[f(x|G)]/[f(x|U)]$ which simplifies the simulation logic. Furthermore, since the exponential density has only one parameter, it is only necessary to change the respective means M_G and M_U of the respective densities $f(x|G)$ and $f(x|U)$ to bring the densities closer

[1] $s_{L_N(N'_U)} = \left\{ \dfrac{\sum_{i=1}^{N_S} [L_N(N'_U, i) - \bar{x}_N(N'_U)]^2}{N_S - 1} \right\}^{1/2}$

together or move them farther apart. Also, in all comparisons the losses associated with correct decisions L_{GG} and L_{UU}, are assumed to be zero and the cost of deciding an unacceptable item is good is assumed to be greater than the cost of deciding a good item is unacceptable.

Results from the four sets of comparisons made are shown in Tables 3 through 6. The appropriate mean values and cost values are shown above each table. It is possible to present these results graphically by plotting $\bar{x}_N(N'_U)$ and L_B against EP_G for a given value of P_G. However, this requires five graphs for each table. For illustrative purposes, one such graph is provided in Figure 1. This graph corresponds to the data for $P_G = 0.40$ in Table 6.

Table 3: Comparison 1.
$M_G = 10, M_U = 5, L_{GU} = 15, L_{UG} = 1$

P_G	EP_G	N'_U	$\bar{L}_B$	$\bar{x}_N(N'_U)$	$^sL_N(N'_U)$	$Z_{\bar{L}_B}$	$P[\bar{L}_N(N'_U) < \bar{L}_B]$
	.45	50	0.5400	0.5520	0.0730	+3.67	0.0000
	.50	50	0.5392	0.5520	0.0730	+3.92	0.0000
.55	.55	49	0.5388	0.5486	0.0987	+2.22	0.0132
	.60	49	0.5394	0.5486	0.0987	+2.08	0.0188
	.65	49	0.5418	0.5486	0.0987	+1.54	0.0618
	.55	49	0.6322	0.6439	0.0861	+3.04	0.0012
	.60	49	0.6306	0.6439	0.0861	+3.45	0.0003
.65	.65	49	0.6299	0.6439	0.0861	+3.63	0.0000
	.70	47	0.6312	0.6577	0.1460	+4.06	0.0000
	.75	47	0.6375	0.6577	0.1460	+3.09	0.0010
	.65	49	0.7179	0.7394	0.0763	+6.30	0.0000
	.70	47	0.7144	0.7256	0.1164	+2.15	0.0158
.75	.75	47	0.7125	0.7256	0.1164	+2.57	0.0051
	.80	44	0.7167	0.7418	0.1846	+3.04	0.0012
	.85	41	0.7421	0.7970	0.2484	+4.94	0.0000
	.75	47	0.7875	0.8109	0.0860	+6.08	0.0000
	.80	44	0.7767	0.7861	0.1309	+1.60	0.0548
.85	.85	41	0.7697	0.7786	0.1801	+1.10	0.1367
	.90	37	0.7975	0.8164	0.2382	+1.77	0.0384
	.95	21	1.2142	0.1843	0.4895	- 1.36	0.9131
	.85	41	0.7973	0.8050	0.1026	+1.68	0.0465
.95	.90	37	0.7325	0.7519	0.1335	+3.25	0.0006
	.95	21	0.6492	0.6463	0.2766	- 0.23	0.5910
	1.00	0	0.7500	0.7518	0.4580	+0.09	0.4641

Table 4: Comparison 2
$M_G = 10, M_U = 5; L_{GU} = 2, L_{UG} = 1$

P_G	EP_G	N'_U	$\bar{L}_B$	$\bar{x}_N(N'_U)$	$^sL_N(N'_U)$	$Z_{\bar{L}_B}$	$P[\bar{L}_N(N'_U) < \bar{L}_B]$
	.30	48	0.3709	0.3803	0.0687	+3.06	0.0011
	.35	47	0.3679	0.3756	0.0705	+2.44	0.0073
.40	.40	46	0.3667	0.3718	0.0713	+1.60	0.0548
	.45	45	0.3684	0.3732	0.0723	+1.91	0.0281
	.50	42	0.3750	0.3893	0.0763	+4.19	0.0000
	.40	46	0.4444	0.4559	0.0731	+3.52	0.0000
	.45	45	0.4396	0.4488	0.0738	+2.79	0.0026
.50	.50	42	0.4375	0.4399	0.0785	+0.68	0.2483
	.55	40	0.4406	0.4444	0.0789	+1.08	0.1401
	.60	36	0.4531	0.4680	0.0868	+3.84	0.0000
	.50	42	0.5000	0.5101	0.0716	+3.15	0.0008
	.55	40	0.4914	0.4980	0.0739	+1.99	0.0233
.60	.60	36	0.4875	0.4903	0.0785	+0.80	0.2119
	.65	32	0.4939	0.4961	0.0883	+0.56	0.2877
	.70	25	0.5222	0.5290	0.1019	+1.49	0.0681
	.60	36	0.5219	0.5328	0.0784	+3.11	0.0009
	.65	32	0.5043	0.5144	0.0851	+2.65	0.0030
.70	.70	25	0.4958	0.5023	0.0940	+1.54	0.0618
	.75	14	0.5125	0.5208	0.1116	+1.66	0.0485
	.80	1	0.6000	0.5957	0.1353	- 0.21	0.5832
	.70	25	0.4694	0.4806	0.0801	+3.13	0.0009
.80	.75	14	0.4250	0.4258	0.0912	+0.19	0.4247
	.80	1	0.4000	0.3990	0.1074	- 0.21	0.5832
	.85	0	0.4000	0.3999	0.1063	- 0.02	0.5080

Table 5: Comparison 3
$M_G = 20, M_U = 2; L_{GU} = 15, L_{UG} = 1$

P_G	EP_G	N'_U	$\bar{L}_B$	$\bar{x}_N(N'_U)$	$s_{L_N}(N'_U)$	$Z_{\bar{L}_B}$	$P[\bar{L}_N(N'_U) < \bar{L}_B]$
	.15	48	0.1900	0.3137	0.0659	+41.97	0.0000
	.25	46	0.1835	0.2737	0.0659	+30.60	0.0000
.35	.35	42	0.1815	0.2302	0.1225	+ 8.89	0.0000
	.45	39	0.1837	0.2575	0.1796	+ 9.06	0.0000
	.55	35	0.1915	0.7404	0.5343	+22.97	0.0000
	.35	42	0.2687	0.3950	0.0695	+40.63	0.0000
	.45	39	0.2624	0.3389	0.0769	+22.24	0.0000
.55	.55	35	0.2599	0.2998	0.1274	+ 7.00	0.0000
	.65	32	0.2637	0.3403	0.2197	+ 7.80	0.0000
	.75	26	0.2813	0.8276	0.5511	+22.16	0.0000
	.55	35	0.3284	0.4528	0.0648	+42.92	0.0000
	.65	32	0.3181	0.3960	0.0718	+24.25	0.0000
.75	.75	26	0.3130	0.3323	0.1421	+ 3.04	0.0012
	.85	21	0.3273	0.4608	0.3238	+ 9.22	0.0000
	.95	12	0.5314	1.3918	0.6849	+28.09	0.0000
	.65	32	0.3453	0.4920	0.0552	+59.42	0.0000
	.75	26	0.3288	0.3906	0.0909	+15.20	0.0000
.85	.85	21	0.3184	0.3622	0.1683	+ 5.82	0.0000
	.95	12	0.4009	0.6801	0.4573	+14.65	0.0000
	.75	26	0.3446	0.4711	0.0304	+93.05	0.0000
	.85	21	0.3095	0.3820	0.0633	+25.61	0.0000
.95	.90	18	0.2880	0.3431	0.0957	+12.87	0.0000
	.95	12	0.2704	0.2909	0.1658	+ 2.76	0.0029
	1.00	0	0.7500	0.7338	0.4561	- 0.79	0.7852

Table 6. Comparison 4
$M_G = 20, M_U = 2; L_{GU} = 2, L_{UG} = 1$

P_G	EP_G	N'_U	$\bar{L}_B$	$\bar{x}_N(N'_U)$	$s_{L_N}(N'_U)$	$Z_{\bar{L}_B}$	$P[\bar{L}_N(N'_U) < \bar{L}_B]$
	.10	47	0.0927	0.1413	0.0547	+19.87	0.0000
.20	.20	44	0.0894	0.1104	0.0506	+ 9.29	0.0000
	.30	40	0.0919	0.1594	0.0693	+21.78	0.0000
	.40	35	0.0995	0.3004	0.0956	+46.99	0.0000
	.20	44	0.1634	0.2772	0.0696	+36.56	0.0000
	.30	40	0.1558	0.2077	0.0644	+18.02	0.0000
.40	.40	35	0.1533	0.1785	0.0611	+ 9.22	0.0000
	.50	32	0.1563	0.2057	0.0792	+13.95	0.0000
	.60	26	0.1675	0.3334	0.1113	+33.33	0.0000
	.40	35	0.2071	0.3039	0.0666	+32.50	0.0000
	.50	32	0.1986	0.2554	0.0642	+19.79	0.0000
.60	.60	26	0.1951	0.2130	0.0683	+ 5.86	0.0000
	.70	21	0.2009	0.2499	0.0937	+11.69	0.0000
	.80	14	0.2320	0.3931	0.1203	+29.94	0.0000
	.60	26	0.2226	0.3293	0.0555	+41.90	0.0000
	.70	21	0.2066	0.2540	0.0546	+19.41	0.0000
.80	.80	14	0.1979	0.2088	0.0676	+ 3.61	0.0000
	.90	6	0.2326	0.2818	0.1034	+10.63	0.0000
	1.00	0	0.4000	0.3982	0.1149	- 0.35	0.6368
	.70	21	0.2095	0.3228	0.0424	+59.75	0.0000
.90	.80	14	0.1808	0.2125	0.0479	+14.80	0.0000
	.90	6	0.1588	0.1676	0.0690	+ 2.85	0.0022
	1.00	0	0.2000	0.2058	0.0836	+ 1.55	0.0606

Conclusions

The results indicate that the Bayes procedure consistently yields lower costs than the N-procedure. While these results are expected for the cases in which the a priori probability P_G is known, the behavior of these costs in presence of errors in the estimates of P_G could not have been predicted in advance. A general trend observed in the comparison graphs is that the two cost curves tend to be closest to each other when $EP_G = P_G$, i.e., in general, the probability $P[\bar{L}_N(N'_U) < \bar{L}_B]$ decreases as the error $|P_G - EP_G|$ increases. There are exceptions to this generalization; but they only occur when data variability is very high and, as such, can be reasonably attributed to random variation.

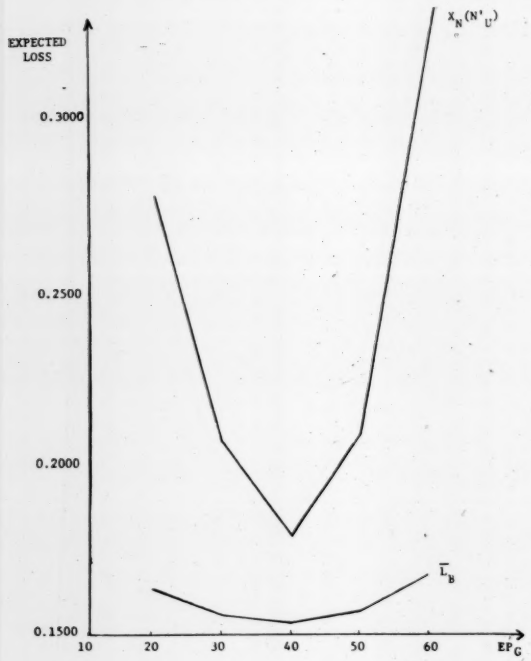


Fig. 1. $M_G = 20$; $M_U = 2$; $L_{GU} = 2$; $L_{UG} = 1$; $P_G = 0.40$.

Therefore, if the Bayes procedure is to be chosen over the N-procedure when P_G is known, it should be chosen, with even more reason, in the presence of errors in the estimates of P_G. It is also useful to note that as M_G/M_U moves away from 1, the expected costs associated with both procedures decrease (assuming L_{GU} and L_{UG} do not change). This is because for M_G/M_U close to 1, the two density functions are close to each other and the probabilities of incorrect decisions are higher. As the densities are moved apart ($M_G/M_U \gg 1$ or $\ll 1$) these probabilities decrease and so do the corresponding expected costs. It is also observed that the farther the ratio M_G/M_U is from 1, the more improbable it becomes that the N-procedure outperforms the Bayes procedure.

References

[1] Rosenshine, M., "Optimal Inspection Procedures," *AIIE Transactions*, **2** 4, pp. 340-347 (December 1970).

Dr. Matthew Rosenshine is Professor, Department of Industrial and Management Systems Engineering, The Pennsylvania State University. Dr. Rosenshine previously held the position of principal mathematician at the Cornell Aeronautical Laboratory of Cornell University and was concurrently affiliated with the State University of New York at Buffalo. He has also worked at the Bell Aircraft Corporation and held part-time teaching appointments at New Mexico State University, University of Alabama, and University of Buffalo. He holds an AB in Science and a MA in Mathematics Education from Columbia University, and a MS in Mathematics from the University of Illinois, a PhD in Operations Research Engineering from The State University of New York at Buffalo. Dr. Rosenshine is a member of the Editorial Board of *AIIE Transactions* and a Senior Member of AIIE. Other society memberships include ORSA, TIMS, AIDS, AAAS, Sigma Xi, Phi Beta kappa, and Alpha Pi Mu.

Francisco Soares has spent the last four years with the General Electric Company. He is currently in charge of the Manufacturing Engineering Group for Locomotion and Traction equipment at the General Electric Company plant in Sao Paulo Brazil. In addition, he is a part-time instructor of Engineering Economy and has taught courses in General Electric's Development Program for Engineers. He holds an MS Degree in Industrial Engineering from The Pennsylvania State University, and undergraduate degrees in Electrical Engineering and Business Administration from The Federal University of Pernambuco in Brazil.

An Improved Algorithm for Scheduling Jobs on Identical Machines

J. WESLEY BARNES
SENIOR MEMBER, AIIE
The University of Texas at Austin

J. J. BRENNAN
Texas Petroleum Research Committee
The University of Texas at Austin
Austin, TX 78712

Abstract: The problem discussed in this paper is one in which n jobs are to be scheduled on M identical machines. Each job, i, possesses a processing time, t_i, a linear penalty rate, p_i, and a due date, $d_i = t_i$ for all i. The objective is to minimize the total penalty cost,

$$C = \sum_{i=1}^{n} p_i(T_i - t_i)$$

where T_i is the actual completion time for job i. The improved algorithm is based upon the technique of Elmaghraby and Park and is the result of three theoretical refinements of their basic method in conjunction with newly developed results. A detailed discussion of the refinements and new results is followed by a presentation of some representative computational results using the improved algorithm.

■ The problem discussed in this paper is one in which n jobs are to be scheduled on M identical machines. Each job, i, possesses a processing time, t_i, a linear penalty rate, p_i, and a due date, $d_i = t_i$ for all i. The objective is to minimize the total penalty cost,

$$C = \sum_{i=1}^{n} p_i(T_i - t_i)$$

where T_i is the actual completion time for job i. Since our results are a direct extension of the basic branch-and-bound algorithm of Elmaghraby and Park [6], we employ consistent notation. All symbols and their meanings are given below for the reader's convenience.

Definitions and Notation

t_i	: processing time for job i
d_i	: due time for job i
p_i	: penalty per unit time of tardiness in job i
r_i	: t_i/p_i, the ratio of processing time to the penalty rate
$Q{:}i,j,\ldots,\ell$	: a sequence in which job i is first, job j next, ..., and job ℓ last
M	: total number of identical machines
Q_m	: $\{m_1, \ldots, m_k\}$ a sequence of k jobs on machine m; m=1, 2, ..., M
m_j	: the jth job on machine m (m_ℓ denotes the last job on machine m)
s_i	: start time of job i
T_i	: time of completion of job i in a given sequence
n	: the size of the set of jobs $N = \{1, 2, \ldots, n\}$
N	: the set of jobs $\{1, 2, \ldots, n\}$
H	: $\sum_{i=1}^{n} t_i/M$

Received February 1976; revised August 1976.

$P(k)$: an eligible permutation of k jobs, i.e., $P(k) = \{J(1), J(2), \ldots, J(k)\}$ where $J(i)$ denotes the job scheduled in position i, $i=1, \ldots, k$. For any $P(k)$ there remains $n-k$ jobs to be scheduled.

$A[P(k)]$: The set of jobs available for scheduling in position $(k+1)$. Note that precedence relationships may allow $A[P(k)]$ to be a set of less than $n-k$ elements.

Ω : the set $\{m_\varrho\}$, $m=1, \ldots, M$. If no jobs have been scheduled on machine m, $m_\varrho \equiv 0$.

λ : $T_\lambda = \min T_{m_\varrho}$, $m_\varrho \in \Omega$

$\bar{P}$: $N - P(k)$

Assumptions

As in [5, 6, 8] we assume:

i) Set up and processing times are independent of the order in which the jobs are scheduled, and the set-up times are included in the processing times.

ii) Each machine begins at time zero and operates without idle time on one job at a time, until all jobs assigned to it are completed.

iii) Each job, once started, must be performed to completion (no job cancellation).

iv) No known precedence relationships exist among the n jobs.

v) The jobs are numbered in "natural order," i.e. in order of non-decreasing r_i with ties broken by giving the job with smaller t_i the smaller number. In the case that $p_i=p_j$ and $t_i=t_j$, ties are broken arbitrarily. Therefore, $i < j$ implies $r_i \leqslant r_j$. (All indexed summations, i.e. Σ_1^k are taken in this order.)

Our purpose is not to reiterate the development of the basic algorithm, and we assume the reader is familiar with the theoretical results presented in [6]. We will, however, discuss refinements of that algorithm and present newly developed results.

Refinements to the Basic Algorithm

Refinement (a) – In the basic algorithm, precedence relations for the n jobs are constructed by using the following results [6, 8]:

Lemma 1. For $i < j$, if $t_i \leqslant t_j$ and $p_i \geqslant p_j$, then job i precedes job j ($i \prec j$) in optimal schedule.

Lemma 2. There exists an optimal schedule in which job 1 is scheduled first.

Lemma 3. In an optimal schedule if $M = 2$, $h \prec k$ if both the following conditions are satisfied:

(i) $r_h < r_{h+1}$, and $r_{k-1} < r_k$

(ii) $$\sum_1^{h-1} t_i \leqslant H - \tfrac{1}{2} \max_{i \in N} t_i - \sum_k^n t_i \quad .$$

Lemma 4. In an optimal schedule on $M \geqslant 3$ machines $h \prec k$ if both of the following conditions are satisfied:

(i) $r_h < r_{h+1}$ and $r_{k-1} < r_k$

(ii) $$\sum_1^{h-1} t_i \leqslant H - \max_{i \in N} t_i - \sum_k^n t_i .$$

Lemma 1 has been restated with the proviso that $i < j$. In the case where $p_i = p_j$ and $t_i = t_j$, a reciprocal precedence relationship would develop were it not for this added restriction.

Condition (i) in Lemmas 3 and 4 is not necessary. Strict adherence to it can cause certain identifiable precedence relationships to be ignored; oversights of this type can have significant effects on the amount of computation needed to achieve and prove optimality.

For example, consider problem 6 of Table 1. Since only two machines are present, Lemmas 1, 2, and 3 yield the following precedence relations:

$$1 \prec 2, 3, 4, 5, 6, 7, 8, 9, 10; 2 \prec 9; \text{and } 3 \prec 9.$$

However, removing condition (i) yields the following additional precedences:

$$2 \prec 10, 3 \prec 10 \text{ and } 4 \prec 10.$$

Let us state a well known result as Lemma 5, present the refinement of Lemma 3 in theorem form and construct its proof.

Lemma 5: At least one optimal schedule Q will have each set of jobs, Q_m, $m=1, \ldots, M$ sequenced in natural order.

Theorem 1: In at least one optimal schedule on two machines (which satisfies Lemma 5), $h \prec k$ if the following condition is satisfied:

$$\sum_1^{h-1} t_i \leqslant H - \tfrac{1}{2} \max_{i \in N} t_i - \sum_k^n t_i .$$

Before proceeding with the proof of Theorem 1, a few

comments should be made. First, we seek an optimal solution satisfying Lemma 5, i.e. a solution with natural order scheduling on each machine. At the outset, the natural order schedule may not be unique, since ties can occur among the r_i's. However, once the jobs have been indexed, we have a unique natural ordering of the jobs. Unlike Lemmas 3 and 4, Theorem 1 is not a qualification for *any* optimal solution, but concerns one optimal solution to which we turn our attention in the algorithm.

Proof: Assume the condition is satisfied *and* $h > k$. Then, by assumption,

$$\sum_{1}^{h-1} t_i \leqslant H - \tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{n} t_i,$$

and $$\sum_{1}^{h-1} t_i \leqslant H - \tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{h-1} t_i - \sum_{k}^{n} t_i.$$

Thus, $$\sum_{1}^{n} t_i = 2H \leqslant H - \tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{h-1} t_i,$$

and $$H \leqslant -\tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{h-1} t_i,$$ which forms a contradiction.

Therefore, if jobs h and k satisfy the condition, $h < k$.

If h and k are scheduled on the same machine, $h \prec k$ by Lemma 5.

In the other case, assume h and k are scheduled on different machines. Without loss of generality, assume h is scheduled on Machine 1 and k on Machine 2. Then

$T_h - t_h = s_h \leqslant \sum_{1}^{h-1} t_i$ on Machine 1, else Lemma 5 is violated. Also,

$T_{2\ell} \geqslant H - \tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i$, thus

$T_k - t_k = s_k \geqslant H - \tfrac{1}{2} \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{2\ell} t_i$ on Machine 2, since

$\sum_{k}^{2\ell} t_i \leqslant \sum_{k}^{n} t_i$, otherwise Lemma 5 is violated.

The conclusion that $s_h \leqslant s_k$ may be motivated exactly as in [8]. Knowledge that $s_h \leqslant s_k$ and $h < k$ joined with the fact that redundant eligible partial permutations need not be considered is sufficient to conclude that $h \prec k$. Q.E.D.

As in [6], this result may be generalized to the case of more than two machines. The cost is a weakening of the condition sufficient for precedence. This result is stated as

Corollary 1: In at least one optimal schedule on $M \geqslant 3$ machines, $h \prec k$ if the following condition is satisfied:

$$\sum_{1}^{h-1} t_i \leqslant H - \operatorname*{Max}_{i \epsilon N} t_i - \sum_{k}^{n} t_i.$$

Refinement (b) – It should also be noted that the superior lower bound developed by Park [8] has been used in the improved algorithm in preference to the lower bound presented in [6]. (See the Appendix for a discussion of the superior lower bound.)

Refinement (c) – Assume an eligible job i has been scheduled in position k. The basic algorithm calls for the generation of $A[P(k)]$. The intent of the definition of $A[P(k)]$ is to provide a recursive definition of a list of eligible successors to any partial permutation $P(k)$. In order to assure that the entire list of eligible successor jobs is maintained, $A[P(k)]$ is defined as:

$$A[P(k)] = A[P(k\text{-}1)] - (i) + \{j | j \epsilon A(i) \text{ and } B(j) \subseteq P(k)\}$$

$$+ \{j | (j < i \text{ and } j \epsilon \bar{P}) \text{ and } B(j) \subseteq P(k)\}$$

where $A(i)$: Set of immediate successors of i
$B(j)$: Set of immediate predecessors of j.

The sets $A(i)$ and $B(j)$ are easily identified from the precedence relations constructed for the problem.

The definition of $A[P(k)]$ as given in [6] and [8] forces the reader to assume that, at any redefinition of $A[P(k)]$, $A[P(k\text{-}1)]$ has been updated to contain any $\{j | (j < i$ and $j \epsilon \bar{P})$ and $B(j) \subseteq P(k)\}$ that may have been deleted at a prior iteration. The above definition provides a clearer picture of the recursive updating of $A[P(k)]$. In addition, the computational results reported in [6] are inconsistent with the stated theory because the updating of $A[P(k)]$ as coded in the PLI program cited in [6] and contained in [8] is in error, [9].

A New Development: Entrapment

The branching rule in the basic algorithm is to branch to the lowest numbered job in $A[P(k)]$, say job i, then i is scheduled on the machine with the earliest available start time, ties being broken by scheduling on the lowest numbered machine. Elmaghraby and Park state they adopted this rule both for simplicity and because the optimal schedule was not expected to be much different from the natural ordering.

Below we present an extension of the basic algorithm which maintains this branching rule but for a much more compelling reason.

Theorem 2: If there exists a job $j \epsilon A[P(k)]$ such that $j < \lambda$ then j is an ineligible first successor to $P(k)$.

Proof: This follows immediately from definition of λ and Lemma 5.

Corollary 2: If there exists a job $j \in A[P(k)]$ such that $j < m_\varrho$, for all $m_\varrho \in \Omega$ then the branches leading from job $J(k-1)$ in $P(k-1)$ have been fathomed. For ease of reference we say j has been trapped, or an entrapment exists.

Proof: We first note that only the lowest job number in $A[P(k)]$ need be considered since if Corollary 2 holds for $i \in A[P(k)]$ and $i > j$, then surely Corollary 2 holds for j. Since $j < m_\varrho$, all $m_\varrho \in \Omega$, the branch leading from j at k in $P(k)$ has been previously fathomed and any other job in $A[P(k)]$ must be $> j$. Thus j is trapped for any other ordering leading from k implying the branch leading from $J(k-1)$ has been fathomed.

We present an example to clarify the above ideas.

Suppose $M = 2$. In Fig. 1, job 7 must have been scheduled at position $(k-1)$ in a previously fathomed permutation. This branch was explored (perhaps leading to a complete schedule). The next available job in $A[P(k)]$ is 8, thus 7 is scheduled at position k as shown in Fig. 1. After the branch

$J(k-2)$	$J(k-1) \equiv 8$	$J(k+1) = \ldots$
$J(k-3)$	$J(k) \equiv 7$	

Fig. 1.

leading from 7 at position k has been fathomed, we schedule the next available job in $A[P(k)]$, say job 9. But now 7 is trapped and both jobs 9 and 8 in Fig. 2 may be unscheduled.*

$J(k-2)$	$J(k-1) \equiv 8$	$J(k+1) = \ldots$
$J(k-3)$	$J(k) \equiv 9$	

Fig. 2.

There are a number of advantages which favor including entrapment in the algorithm.

1) The concept and computations involved are simple and it can cause tremendous savings in search effort.
2) The entrapment method can eliminate non-optimal branches even when the lower bound is ineffective.
3) Entrapment is multi-leveled, i.e., it can eliminate sub-branches at all levels of a main branch.
4) Once entrapment occurs at a node, it has a "snowballing" effect. Referring to Fig. 2, when 8 is eliminated at $J(k-1)$ only a greater numbered job may be scheduled. Then both jobs 7 and 8 are likely candidates for entrapment. The entrapment procedure detects the decreasing likelihood of finding an eligible combination.

*In the example depicted in Fig. 2, we can look at the scheduling of job 9 at k in $P(k)$ as ineligible since it eventually forces the sequence to the condition described by Corollary 2. The algorithm and our coding find entrapment just prior to its occurrence when, in fact, the scheduling of 9 could be considered the entrapment. However, the additional computational effort required for this approach does not seem to be justified.

Computational Results Using the Improved Algorithm

The data and results of fifteen problems solved by the improved algorithm are presented in Table 1. The first ten problems are problems 1 through 10 from [6]. Problem 11 is a "real world" problem that uses actual data from an oil-well workover scheduling project. The optimal sequence from this problem is radically different from "natural" ordering, a condition that did not occur in the work presented in [6]. Again in problem 12, a job is significantly shifted from the natural order in the optimal sequence.

Problem 6 is due to Bhadury [3]. In [6] this problem was attempted using the basic algorithm. 1315 iterations

Table 1

Problem 1		
M	2	Jobs
n	5	1 2 3 4 5 6 7 8 9
Iterations to Best	5	t_i 1 2 3 2 4
Iterations to Prove	8	p_i 5 7 6 3 4
Optimum	26	Optimal Schedule
CPU Time	.36S	1-2-3-4-5
Problem 2		
M	3	Jobs
n	6	t_i 2 3 6 4 4 5
Iterations to Best	6	p_i 8 9 8 5 3 1
Iterations to Prove	10	Optimal Schedule
Optimum	60	1-2-3-4-5-6
CPU Time	.38S	
Problem 3		
M	3	Jobs
n	7	t_i 1 1 4 6 9 4 9
Iterations to Best	19	p_i 5 2 6 8 9 1 1
Iterations to Prove	52	Optimal Schedule
Optimum	26	1-2-4-3-5-6-7
CPU Time	.41S	
Problem 4		
M	2	Jobs
n	8	t_i 3 5 6 8 2 2 7 7
Iterations to Best	11	p_i 7 8 8 7 1 1 3 1
Iterations to Prove	20	Optimal Schedule
Optimum	129	1-2-3-4-5-7-6-8
CPU Time	.46S	

Table 1 (continued)

Problem 5	
M	3
n	9
Iterations to Best	9
Iterations to Prove	22
Optimum	140
CPU Time	.47S

Jobs

t_i	1	4	7	5	5	6	7	7	9
p_i	8	7	8	5	5	4	4	3	3

Optimal Schedule

1-2-3-4-5-6-7-8-9

Problem 6 (Bhadbury's Problem)	
M	2
n	10
Iterations to Best	12
Iterations to Prove	730
Optimum	714
CPU Time	.80S

Jobs

t_i	1	7	6	5	4	11	10	13	3	15
p_i	1	6	5	4	3	8	7	9	2	10

Optimal Schedule

1-2-3-4-5-6-7-8-10-9

Problem 7	
M	3
n	11
Iterations to Best	11
Iterations to Prove	131
Optimum	109
CPU Time	.52S

Jobs

t_i	1	2	1	2	2	1	2	4	7	6	8
p_i	5	9	4	8	7	3	4	7	7	4	2

Optimal Schedule

1-2-3-4-5-6-7-8-9-10-11

Problem 8	
M	2
n	12
Iterations to Best	12
Iterations to Prove	77
Optimum	489
CPU Time	.64S

Jobs

t_i	1	3	4	4	6	4	9	5	8	8	9	9
p_i	5	8	8	6	8	4	6	3	4	4	4	2

Optimal Schedule

1-2-3-4-5-6-7-8-9-10-11-12

Problem 9	
M	3
n	13
Iterations to Best	119
Iterations to Prove	633
Optimum	230
CPU Time	.962S

Jobs

t_i	1	1	3	5	3	6	2	5	8	5	6	7	7
p_i	8	5	9	7	4	8	2	5	8	4	3	2	1

Optimal Schedule

1-2-3-4-6-5-7-9-8-10-11-12-13

Problem 10	
M	3
n	25
Iterations to Best	71
Iterations to Prove	743
Optimum	950
CPU Time	1.69S

Jobs

t_i	1	1	1	2	2	4	4	5	6	6	6	6	4	4	4	7	7	7							
p_i	9	8	8	8	8	8	8	7	8	7	6	6	3	3	5	3	5	3							
t_i	5	6	7	8	8	8	9																		
p_i	2	2	2	2	2	1	1																		

Optimal Schedule

1-2-3-4-5-6-7-8-9-10-11-12-13-16-14-15-17-18-19-20-21-22-23-24-25

Problem 11	
M	5
n	9
Iterations to Best	33
Iterations to Prove	50
Optimum	856
CPU Time	.36S

Jobs

t_i	1	5	6	5	5	11	5	7	16
p_i	57	102	112	88	51	108	40	51	106

Optimal Schedule

1-2-3-4-6-9-5-8-7

Table 1 (continued)

Problem 12	
M	2
n	5
Iterations to Best	13
Iterations to Prove	18
Optimum	38
CPU Time	.31S

Jobs

t_i	1	2	3	2	100
p_i	5	7	6	3	11

Optimal Schedule

1-2-5-3-4

Problem 13	
M	3
n	15
Iterations to Best	17
Iterations to Prove	16585
Optimum	9913
CPU Time	14.63S

Jobs

t_i	2	4	5	5	7	4	4	11	11					
p_i	46	86	93	82	96	52	48	109	100					
t_i	9	5	8	12	8	9								
p_i	77	39	55	78	40	45								

Optimal Schedule

1-2-3-4-5-6-7-8-9-10-11-12-13-15-14

Problem 14	
M	3
n	20
Iterations to Best	22
Iterations to Prove	68334
Optimum	519
CPU Time	84.60S

Jobs

t_i	2	1	8	3	3	8	5	3	7	4	9	5								
p_i	20	2	10	3	2	4	2	1	2	1	2	1								
t_i	12	21	16	9	11	24	13	30												
p_i	2	3	2	1	1	2	1	2												

Optimal Schedule

1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-16-17-18-20-19

Problem 15	
M	4
n	20
Iterations to Best	30767
Iterations to Prove	Search Incomplete
Best Obtained	8755
CPU Time	148.71S

Jobs

t_i	2	2	3	3	2	3	2	6												
p_i	97	84	106	106	67	76	42	101												
t_i	5	8	7	6	10	11	9	6	11											
p_i	70	98	62	47	78	79	63	39	58											
t_i	7	8	11																	
p_i	36	41	32																	

Optimal Schedule

1-2-3-4-5-6-8-10-7-9-11-13-14-12-15-16-17-19-18-20

were reported to be required to prove optimality. After assuring correct updating of $A[P(k)]$, the authors attempted this problem using the basic algorithm, both with Park's lower bound [8] and with the lower bound reported in [6]. In both cases, the number of iterations required to prove optimality was at least an order of magnitude greater than that reported in [6] and [8]. Our own results, using the improved algorithm, are shown in Table 1. The three additional precedence relations generated by the removal of condition (i) in Lemma 3 decreased the number of iterations required to prove optimality with the improved algorithm from 829 to 730, a reduction of 13%.

Problems 13, 14, and 15 are included to indicate the rapidly increasing computational burden that is encountered as the number of jobs increases. A comparison between problem 10 and problems 13, 14, and 15 indicates

that the algorithm's ability to solve larger problems to optimality is highly dependent on individual problem characteristics. Note that the best solution for problem 15 is obtained deep within the search tree.

It has been the authors' experience [4] that for $n > 20$ using actual and simulated data for backlogs of oil well workovers, the optimal solutions generally cannot be proven in less than 150 seconds of CPU time. However, optimal or near-optimal schedules can usually be obtained using a combination of heuristic starting solutions [1] and branch and bound with a limited search [2].

Conclusions

In this paper we have indicated several points where refinements to the basic algorithm of Elmaghraby and Park [6] can be made and have developed a new technique of fathoming branches of the search tree through ineligibility by a process we call "entrapment." Entrapment has been shown to be an effective method of reducing the computational effort in obtaining and proving optimality. However, as is indicated in problems 13, 14, and 15 of Table 1, the size of the problem that can be solved even by this improved algorithm is, at present, more limited than was heretofore believed.

Acknowledgment

The authors wish to thank Prof. Sung H. Park for providing the first ten test problems from [6] for use in the computational study of the improved algorithm.

References

[1] Baker, K. R., and Merten, A. G., "Scheduling with Parallel Processors and Linear Delay Costs," *NRLQ*, **20** 4, pp. 793-804.

[2] Barnes, J. W., Brennan, J. J., and Knapp, R. M., "Optimal Scheduling of a Backlog of Oil Well Workovers," Report No. UT 76-3, Texas Petroleum Research Committee, Submitted to *The Journal of Petroleum Technology*.

[3] Bhadury, B., "An Experimental Investigation of a Model for Sequencing n Jobs on m machines in Parallel," Unpublished thesis, North Carolina State University at Raleigh (1970).

[4] Brennan, J. J., *An Algorithm for Optimally Scheduling* n *Jobs on* M *Machines,* Unpublished Master's Thesis, The University of Texas at Austin (1976).

[5] Eastman, S., Even, S., and Isaacs, I., "Bounds for the Optimal Scheduling of n Jobs on m Processors," *Management Science,* **11** 2, pp. 268-279 (November 1964).

[6] Elmaghraby, S. E., and Park, S. H., "Scheduling Jobs on a Number of Identical Machines," *AIIE Transactions,* **6** 1, pp. 1-13.

[7] McNaughton, R., "Scheduling with Deadlines and Loss Functions," *Management Science,* **6** 1, pp. 1-12 (October 1959).

[8] Park, S. H., "A Branch-and-Bound Method for the Sequencing Problems of n Jobs on m Identical Machines in Parallel," Unpublished Master's thesis, North Carolina State University at Raleigh (1972).

[9] Park, S. H., Personal communication (July 27, 1976).

Appendix

The lower bound given in [6] as

$$LB_1 = \sum_{j \epsilon p(k)} p_j(T_j - d_j) + \text{Max}\left\{0, \frac{1}{M} c_{1,n-k}(T_\lambda) + \frac{M-1}{M} T_\lambda \cdot \sum_{j \epsilon \bar{p}(k)} p_j - \frac{M-1}{2M} c_{n-k,n-k}(0)\right\}$$

may be rewritten as

$$LB_1 = \sum_{j \epsilon p(k)} p_j(T_j - d_j) + \text{Max}\left\{0, \frac{\sum_{j'=1}^{n-k} p_{j'} T_\lambda + \sum_{i'=1}^{j'-1} t_{i'}}{M} + \frac{M-1}{M} T_\lambda \sum_{j'=1}^{n-k} p_{j'} - \frac{M-1}{2M} \sum_{j'=1}^{n-k} p_{j'} t_{j'}\right\}$$

$$LB_1 = \sum_{j \epsilon p(k)} p_j(T_j - d_j) + \text{Max}\left\{0, \frac{T_\lambda \sum_{j'=1}^{n-k} p_{j'}}{M} + \frac{\sum_{j'=2}^{n-k} p_{j'} \sum_{i'=1}^{j'-1} t_{i'}}{M} + \frac{M-1}{M} T_\lambda \sum_{j'=1}^{n-k} p_{j'} - \frac{M-1}{2M} \sum_{j'=1}^{n-k} p_{j'} t_{j'}\right\}$$

$$= \sum_{j \epsilon p(k)} p_j(T_j - d_j) + \text{Max}\left\{0, T_\lambda \sum_{j'=1}^{n-k} p_{j'} + \frac{\sum_{j'=2}^{n-k} p_{j'} \sum_{i'=1}^{j'-1} t_{i'}}{M} - \frac{M-1}{2M} \sum_{j'=1}^{n-k} p_{j'} t_{j'}\right\}$$

where $\bar{p}(k) = N - p(k)$ and j' = the set of job numbers $\{j \epsilon \bar{p}(k)\}$ renumbered into natural order, i.e. if $\bar{p}(k) = \{3, 6, 7\}$, $1' = 3$, $2' = 6$, and $3' = 7$.

The superior lower bound given in [8] as

$$LB_2 = \sum_{j\epsilon p(k)} p_j(T_j - d_j) + T_\lambda \sum_{j\epsilon \bar{p}(k)} p_j + \text{Max} \left\{0, \frac{c_{1,n-k}}{M} - p_{n-k}\frac{M-1}{2M}\right\}$$

may be rewritten as

$$LB_2 = \sum_{j\epsilon p(k)} p_j(T_j - d_j) + T_\lambda \sum_{j'=1}^{n-k} p_{j'} + \text{Max}\left\{0, \frac{\sum_{j'=2}^{n-k} p_{j'} \sum_{i'=1}^{j'-1} t_{i'}}{M} - \frac{M-1}{2M}\sum_{j'=1}^{n-k} p_{j'}t_{j'}\right\}$$

In comparing LB_1 and LB_2, we note the only difference is that the common term $T_\lambda \Sigma_{j'=1}^{n-k} p_{j'}$ is under the "max" operator in LB_1 and outside of the "max" operator in LB_2. Thus LB_2 is always greater than or equal to LB_1 and $LB_2 > LB_1$ when the second argument under the max operator for LB_2 becomes negative.

Dr. J. Wesley Barnes is an Assistant Professor of Mechanical Engineering at the University of Texas at Austin. He holds a BSME from the University of Texas at Austin, and MSIE and PhD in IEOR from the University of Arkansas. Before joining the University of Texas, he was an Operations Research Analyst for the Bell Telephone Laboratories. His current research interests include the scheduling of maintenance functions in the oil industry, the analysis of stochastic cash flows, stochastic linear programming, and applied statistics. He is a member of AIIE, ORSA, TIMS, and a member of the Editorial Board of the *AIIE Transactions*.

Mr. J. J. Brennan is a graduate student working on his PhD at the Graduate School of Management, University of California at Los Angeles. He holds a BA from the University of Pennsylvania and an MSOR&IE from the University of Texas at Austin. He is a member of ORSA, TIMS, and Phi Kappa Phi. This paper was based on Mr. Brennan's Master's Thesis under the direction of Dr. Barnes.

Computational Experience with a Sequencing Algorithm Adapted to the Tardiness Problem

KENNETH R. BAKER
SENIOR MEMBER, AIIE
Graduate School of Business Administration
Duke University
Durham, NC 27706

Abstract: This paper summarizes computational experience with a generalized dynamic programming algorithm known as the Chain Algorithm. This algorithm is a general purpose procedure for solving sequencing problems, and in this study the procedure is adapted to the problem of minimizing total tardiness in the single-machine model. The results indicate that at the present state of the art, 50-job problems can normally be solved in a few seconds of computer time, although memory requirements may be extensive.

■ This paper reports some recent computational experience with the single-machine tardiness problem. This sequencing problem has been the focus of considerable interest among researchers in several areas, partly because of its fundamental importance in scheduling, but also because the techniques developed for it have often illustrated general principles applicable to other combinatorial optimization problems. The latter theme is important in this paper, which deals with the use of a general solution algorithm adapted to the tardiness problem.

The computational results reported here are based on the dynamic programming approach to sequencing problems [see Reference 1] with which the reader is assumed to be familiar. Sequencing problems containing chain structures are amenable to solution by the Chain Algorithm of dynamic programming, as described in general terms in [2]. Here we examine the specific form of the Chain Algorithm when it is adapted to solving the tardiness problem, and we compare its performance to that of other tardiness-oriented algorithms.

In the tardiness problem a set of n jobs is given, and for each job j a processing time (t_j) and a due date (d_j) are specified. A sequence of the jobs will determine the completion time (C_j) for job j, and the job's tardiness is defined as $\max\{0, C_j - d_j\}$. The objective is to find a sequence that minimizes total tardiness in the job set.

The most recent comparative study of tardiness algorithms [1] indicated that with third-generation computers, optimizing algorithms for tardiness problems containing more than 15 jobs tended to consume substantial resources, in terms of CPU time and memory. The most efficient algorithms appeared to be those that exploited the special structure of the tardiness problem in conjunction with general-purpose combinatorial optimization procedures, such as branch and bound or dynamic programming. The major development since that time involves the solution approach of Fisher [4] based on the application of duality-based methods to combinatorial problems. Fisher's results with test problems suggest that 50-job problems can be solved in roughly one minute of CPU time per problem. In the current study, we observe computation times that are considerably faster than those reported previously. We also find that the chain algorithm's effectiveness in solving large problems is ultimately limited by its consumption of core storage, although most tardiness problems containing up to 50 jobs can be solved in only a few seconds of CPU time.

Solution Algorithm

When adapted to the tardiness problem, the Chain Algorithm

Received December 1975; revised June 1976.

consists of three principal modules: Dominance, Chain Construction, and Optimization. In this section we examine these modules in detail.

Dominance

The first step is to check Emmons' dominance conditions [3]. Let the ρ-relation between a pair of jobs be defined as follows:

$i \rho j$ if (1) $t_i < t_j$ or (2) $t_i = t_j$ and $d_i \leqslant d_j$.

Furthermore, let

p_{ij} = 1 if job i dominates job j (i.e., it is sufficient, when seeking an optimum, to consider only sequences in which i precedes j) and $p_{ij} = 0$ otherwise

$A_i = \{j \mid p_{ij} = 1\}$ $A_i' = \{j \mid p_{ij} = 0\}$ $B_j = \{i \mid p_{ij} = 1\}$.

The dominance conditions are:

I. If $i \rho j$ and $d_i \leqslant \max\left\{d_j, t_j + \sum_{k \epsilon B_j} t_k\right\}$ then $p_{ij}=1$.

II. If $i \rho j$, $d_i > \max\left\{d_j, t_j + \sum_{k \epsilon B_j} t_k\right\}$, and $d_i + t_i \geqslant \sum_{k \epsilon A_j'} t_k$, then $p_{ji} = 1$.

III. If $i \rho j$ and $d_j \geqslant \sum_{k \epsilon A_i'} t_k$ then $p_{ij} = 1$.

The dominance conditions are checked recursively (see [3]) in order to construct the matrix of 0-1 values of p_{ij}.

As a result of the information thus gathered, it may be possible to assign a particular sequence position to certain jobs. For instance, if $|B_i| = n-1$ then job i may be placed last in sequence, or if $|A_j| = n-1$ then job j may be placed first in sequence. In either case, the problem size (n) is reduced by one, and the remaining jobs may be checked to see whether further assignments are possible. In some problems it is possible to construct an entire optimal sequence at this step; otherwise, it is necessary to proceed to controlled enumeration. Experiments with randomly-generated problems [1, 4] suggest that the dominance conditions reduce the original problem size by an average of roughly 50%, although in some cases they achieve no reduction at all.

Chain Construction

The information in the P-matrix is utilized by a heuristic procedure for chain construction. A chain is a set of r jobs containing $r-1$ precedence restrictions such that each job in the chain has at most one direct successor and at most one direct predecessor. The heuristic procedure attempts to select precedence relations from the P-matrix in order to construct as many large chains as possible. Specifically, if n_i (i=1, 2, . . . , m) is the number of jobs in chain i, the objective is to minimize $Z = \prod_{i=1}^{m} (n_i + 1)$. As we shall see below, the quantity Z is a measure of both the computational effort and the storage consumption required by the algorithm.

The heuristic procedure finds the job j for which $|A_j|$ is largest. Among jobs $k \epsilon A_j$, the one with the largest $|A_k|$ is selected to be the direct successor of j. A chain is constructed in this fashion by continuing until the procedure reaches a dead end. The jobs in the chain are then removed from consideration and the process is repeated until all jobs are assigned to chains.

The heuristic procedure may not minimize Z, and other procedures could be used (including a symmetric procedure which selects direct predecessors by finding the largest $|B_k|$ among $k \epsilon B_j$). The results to be reported below, however, are based on the single, one-pass procedure described above.

Optimization

The optimal tardiness is found by means of the task-chain version of dynamic programming [2]. The elements of the recursion are as follows. Let:

I = the set of all n jobs

J = a subset of I

$q_J = \sum_{k \epsilon J} t_k$ = the time required to process the jobs contained in J.

(i, k) = the k*th* job in chain i.

k_i^* = the job in chain i most recently added to a given subset J.

The recursion begins with $G(\phi) = 0$ and ultimately determines $G(I)$, where

$$G(J) = \min_{(i,k_i^*) \epsilon J} \left\{\max(0, q_J - d_{i,k_i^*}) + G\left(J - [(i, k_i^*)]\right)\right\}. \quad (1)$$

Notice that this form represents a forward recursion, contrary to the backward recursion that has historically been advocated in the treatment of dynamic programming methods in sequencing. The forward algorithm can exploit a computational "trap door" based on the following property:

If there exists a job $(i, k_i^*) \epsilon J$ such that $d_{i,k_i^*} \geqslant q_J$ then (1) simplifies to $G(J) = G(J - \{(i, k_i^*)\})$.

In any event, the recursion must systematically evaluate the minimum in (1) for each of $Z-1$ subsets in order to find the minimum tardiness and the sequence that achieves it.

Computational Results

The chain algorithm was first tested on the 15-job problem sets used previously in [1]. CPU times[1] were compared to those previously observed for conventional dynamic programming and for Srinivasan's hybrid algorithm, [5] which was the most efficient of the algorithms explicitly compared in [1]. The computation times for the chain algorithm were observed on an IBM 370/165, which is considered to be roughly five times faster on these problems than the computer used in the comparative study.

The storage requirements of the conventional dynamic programming algorithm depend on whether auxiliary (low-speed) memory is utilized. If no auxiliary memory is used then the algorithm requires core storage space for at least one main vector of length 2^n-1, or 32767 in a 15-job problem. A second vector of the same length can be used to efficiently reconstruct the optimal sequence; however, even if this vector is not maintained the additional time spent in the reconstruction phase will be negligible in large problems. The core requirement could be reduced to one main vector of size 6435 by extensive use of auxiliary memory and a concomitant sacrifice in computation time due to the additional overhead involved. However, the gains achieved by storage-efficient coding will not provide substantial relief in large problems,[2] hence the present study used time-efficient coding strategies.

In the hybrid algorithm the dominance relations are used to fill sequence positions where possible (as described in the previous section) and then to selectively avoid dynamic programming calculations where possible. For each calculation to be avoided, however, certain special conditions must be verified; by contrast, the chain algorithm is streamlined once the chains are constructed. The storage requirements of the hybrid algorithm depend on the data in a given problem. The comparative study accentuated the computational time efficiencies of the hybrid algorithm by reserving fully as much core storage as dynamic programming requires, and this figure is given in Table 1. In principle, the hybrid algorithm could be accommodated in far less storage space, since efficient use can be made of the data in the *P*-matrix. On the other hand, computation times for such a version would be slower than those in the table, and there would be no simple way to know in advance the precise storage requirement in a given problem.

The storage requirements of the chain algorithm are greater than those theoretically achievable with the hybrid algorithm, due to the fact that selective use of the information in the *P*-matrix is made. However, the core requirement

Table 1. A comparison of the chain algorithm with other procedures, for the 15-job test problems used in [1].

	Average CPU time (seconds)	Maximum CPU time (seconds)	Main Vector Length
Conventional D.P.	52.6*	52.6*	32767
Hybrid Algorithm	1.13*	7.15*	32767
Chain Algorithm	0.02§	0.04§	1007

*IBM 360/67
§IBM 370/165

in a given problem can be anticipated in advance, since at least one vector of length (Z-1) will be needed when no auxiliary storage is utilized. The figure shown in the table (1007) is the maximum value of (Z-1) observed in the data set; the average value was 174.

The second phase of the experiments involved a comparison with the dual algorithm on the larger job sets used by Fisher [4]. All test problems containing 20 or 30 jobs, and all but one of the problems containing 50 jobs, were solved by the chain algorithm. The solution times in these problems were quite fast, as shown in Table 2.

Table 2: A comparison of the dual and chain algorithms for Fisher's test problems.

	Average CPU time (seconds)	Maximum CPU time (seconds)
25 20-job problems		
Dual Algorithm*	2.82	10.19
Chain Algorithm§	0.03	0.13
25 30-job problems		
Dual Algorithm*	10.08	58.75
Chain Algorithm§	0.15	0.88
24 50-job problems		
Dual Algorithm*	49.40	166.84
Chain Algorithm§	3.54	15.71

*IBM 360/67
§IBM 370/165

Although the chain algorithm exhibited solution time performance that was substantially better than the dual algorithm in these cases, the chain algorithm is ultimately subject to the same type of limitation as conventional dynamic programming: in large problems its core requirements become overwhelming. The approach of the chain algorithm provides insulation from this effect for problem sizes of up to 30 jobs. In these test problems, the maximum value of Z-1 was 21599. (This number is less than the size of the largest vector that would be required by the conventional approach, in its core-efficient form, to solve a 17-job problem.) However, in the 50-job problems the core requirement was sometimes extensive, and one of Fisher's problems required too much memory for a solution even to be

[1] All times reported in this paper are exclusive of I/O and represent CPU times observed under Fortran G.

[2] For example, suppose that the size of core memory limited the size of the main vector to 66000, thus precluding the time-efficient version of the algorithm from solving problems with more than 16 jobs. Even if the storage-efficient coding strategy were employed, problems containing more than 18 jobs would still be too large for the core available.

attempted.[3] The detailed data in Table 3 provide some additional perspective on the core requirements encountered.

Table 3: Details of observations for Fisher's 50-job problems.

Range of Main Vector Size	Number of Problems	Average (Maximum) CPU seconds	
		Chain Algorithms §	Dual Algorithm*
0-9743	14	0.21(0.45)	20.83(139.51)
10000-73727	5	3.57(5.10)	68.04(123.50)
100000-272159	5	12.83(15.71)	110.76(166.84)
1693439	1	-	401.62
		§IBM 360/165	*IBM 370/67

Conclusions

The computational experience reported here indicates that the chain algorithm provides the fastest solution times yet observed for the single-machine tardiness problem. Along with solution time efficiency, however, the algorithm is characterized by a substantial core requirement, and in large problems this factor might render the algorithm impractical. The experiments showed that the chain algorithm's approach was insulated from the demand for prohibitive amounts of core for problem sizes of up to 30 jobs, but the core requirement of some 50-job problems was extensive. We must remember, however, that this requirement is partly a function of the method of chain construction which comprises the second module of the algorithm. We used only a simple heuristic method in this module, whereas a more sophisticated approach could provide some relief from the core requirements we observed. Thus one possible area for further investigation is an efficient procedure for chain construction. A related topic is an extension of the approach to precedence structures more general than chains. Further research along these lines may well lead to significant developments in the use of dynamic programming to solve sequencing problems.

Acknowledgment

Some very helpful advice on tactics for efficient coding of the chain algorithm was provided by Don Erlenkotter of the Graduate School of Management at UCLA and by Linus Schrage of the Graduate School of Business at the University of Chicago.

References

[1] Baker, K. R. and Martin, J. B., "An Experimental Comparison of Solution Algorithms for the Single Machine Tardiness Problem," *Naval Research Logistics Quarterly* (January 1974).

[2] Baker, K. R., and Schrage, L. E., "Finding an Optimal Sequence by Dynamic Programming: An Extension to Task Chains," ORSA-TIMS Special Interest Conference on the Theory and Applications of Scheduling, Orlando (February, 1976).

[3] Emmons, H., "One Machine Sequencing to Minimize Certain Functions of Job Tardiness," *Operations Research* (July 1969).

[4] Fisher, M. L., "A Dual Algorithm for the One-Machine Scheduling Problem," *Mathematical Programming* (to appear).

[5] Srinivasan, V. "A Hybrid Approach to the One Machine Sequencing Problem to Minimize Total Tardiness," *Naval Research Logistics Quarterly* (September 1971).

Dr. Kenneth R. Baker is Associate Professor in the Graduate School of Business Administration at Duke University, where he teaches operations management and management science. He is a Senior Member of AIIE and also a member of TIMS, AIDS, ORSA, and AAUP.

[3] Using a simple linear regression model in which CPU times are a function of Z, we predict that the solution time for the prohibitively large problem would be about 1 minute, if the requisite storage capacity were available.

A Scheduling Algorithm for the Unbalanced Production Line

RUSSELL E. JACOBSON
SENIOR MEMBER, AIIE
University of Wisconsin
Whitewater, Wisconsin

RANDALL P. SADOWSKI
SENIOR MEMBER, AIIE
Purdue University
West Lafayette, Indiana 47907

Abstract: This paper presents procedures for obtaining feasible fractional and integer assignments which can be used to schedule manpower for the unbalanced production line. The objective is to minimize the cost of the total in-process inventory. A mathematical model is formulated and then used to calculate fractional assignments using linear programming and integer assignments using integer programming. An algorithm is then developed which provides the equivalent fractional solution as obtained from linear programming followed by a heuristic algorithm which converts these fractional assignments to feasible integer assignments. Sample problems and their solutions are provided to illustrate the assignment procedures.

■ A production line is generally characterized by a sequence of fixed operations that do not lend themselves to divisibility into small tasks that can be moved to a preceding or succeeding operation. Therefore, the total time required to perform each operation in the sequence is, in general, different than the time required for any of the other operations. Furthermore, the operation times will generally not be multiples of one another. Thus, it is impossible to balance the line, even by assigning varying number of workers to the various operations. In-process inventories between each pair of operations in the sequence are inevitable on a production line, since any given operation will, in general, be producing at a rate that differs from its immediate predecessor(s) and its immediate successor. Further, the quantities in these in-process inventories will be constantly changing with time.

Control of a production line is accomplished through flow control, or control of the rate of flow of production from one work station to another. Assuming that the required number of work stations is available and that operation times are deterministic and unbalanced, flow control is achieved through inventory control. If the required raw materials are available, inventory control for the work-in-process and finished product is achieved through scheduling. If all manpower assigned to the production line is fully utilized, usage of machine and material can be considered as being dependent on the assignment of manpower. Therefore, the problem of controlling a production line can be reduced to that of attempting to schedule the manpower in such a way as to minimize some measure of the work-in-process inventory and maximize the output of the finished product, on a period to period basis. At present most industrial firms tend to control the in-process inventory on their unbalanced production line(s) with some variation of the minimum-maximum inventory control technique. When some minimum value of a particular type of in-process inventory is reached, an additional worker is assigned to the operation that produces that particular type of in-process inventory. Similarly, when a maximum value is reached, a worker is removed from the appropriate operation.

The implementation of minimum-maximum inventory control techniques creates many problems. Valuable time is wasted by the repeated monitoring of quantities of each type of in-process inventory. The method of reassignment, as the result of reaching a minimum or maximum, often creates a need for another reassignment within a short time. Delays in shifting workers can cause variations in the output of the line and the frequent shifting of workers causes labor to be less than fully utilized. Overtime and undertime must frequently be used in attempt to bring various in-process inventories to the desired levels and the cost of holding in-process inventory tends to be excessive.

Received April 1976; revised September 1976.

A great deal of research has been conducted towards a solution of problems in scheduling operations and controlling inventory levels. Much of the work published prior to 1974 has been collected in a 1974 monograph edited by Montgomery and Berry [6] that contains articles that summarize research in specific areas. The references listed at the end of each summary article, coupled with a guide to the literature at the end of the monograph, emphasize the great amount of effort that researchers have expended in seeking solutions to these problems. Additional research results, published both prior and subsequent to the publication of this monograph, indicate a similar effort and diversity of techniques.

Although a great deal of research has been conducted in the general area of inventory control, very little can be found which relates directly to the scheduling of manpower and/or controlling of in-process inventories for the unbalanced production line. The most recent work that considers the in-process inventory problem directly makes use of various techniques: a finite Markov chain [5], dynamic programming [4], computer simulation [1], and an alternative to linear programming [2, 3]. Davis [3] presented a technique which controls in-process inventory and "balances" the production line by assigning a fractional number of workers to each operation during each time period. This technique cannot be used directly, since it is not possible to make fractional manpower assignments in the real world. In summary, none of the techniques found in the literature are directly applicable to the problem of production line scheduling with integer manpower assignments so as to optimize in-process inventory and line output on a day-to-day basis.

Model Formulation

It is possible to apply classic optimization techniques to create an optimizing model for the general production line problem. The basic assumptions used in creating the model are as follows:

- –the rates of output for the operations are deterministic;
- –the costs of holding the in-process inventory are known and deterministic;
- –the quantity of in-process inventory between each sequence of operations is assumed to be known at the start of each scheduling period;
- –the total number of workers in any time period is known, but may vary period-to-period;
- –each worker can perform all the operations, but must be assigned to one operation for the full scheduling period;
- –the required number of work stations of each type are available;
- –each worker assigned to the line is to be fully utilized for the entire period.

The final operation in the production line segment under consideration is considered to have the same defining variables as the other operations in the line. Thus, the option exists to handle the in-process inventory following the final operation as input for either of two operations: finished goods warehouse or another segment of the total production line.

Each operation on the production line segment under consideration is assigned a positive integer j as a means of identification. A part that has been processed through operation j and is awaiting processing at the direct successor of operation j is identified as part j.

The following variables are used to describe the production line:

NW total number of workers on the line;

NO total number of operations on the line;

$LPN(j)$ lowest numerical designation of any of the direct predecessors of operation j, $(j = 1, 2, \ldots, NO)$;

$NDP(j)$ total number of direct predecessors of operation j, $(j = 1, 2, \ldots, NO)$;

$NDS(j)$ numerical designation of the direct successor of operation j, $(j = 1, 2, \ldots, NO)$;

$N(j)$ number of workers assigned to operation j, $(j = 1, 2, \ldots, NO)$;

$$\sum_{j=1}^{NO} N(j) = NW;$$

$A(j)$ average output of good parts from operation j per worker per time period, $(j = 1, 2, \ldots, NO)$;

$C(j)$ average cost of holding one unit of part j as in-process inventory per time period, $(j = 1, 2, \ldots, NO)$;

$U(j)$ usage of units of part j per good part produced at operation $NDS(j)$, $(j = 1, 2, \ldots, NO)$;

$E(j)$ number of units contained in the in-process inventory that exists between operation j and operation $NDS(j)$ $(j = 1, 2, \ldots, NO)$.

The value of $C(NO)$ will be zero in many cases, since it will often be desirable to maximize the output of finished goods. However, the option exists to set the value of $C(NO)$ to any value to reflect the actual inventory cost structure. The output of finished goods from the production line segment under consideration can be obtained from the product of $N(NO)$ and $A(NO)$. It may initially appear that the three vectors used to define the line configuration, LPN, NDP, and NDS, are unnecessarily awkward as mathematically they could be easily replaced with a single vector. However, the subsequent models were developed for use by production personnel and this notation was found to simplify the implementation of the models.

The general, non-trivial production line operating under the assumptions listed previously requires that:

$NW \geq 1$;

$NO \geq 2$;

$LPN(j)$ { = -1, if operation j has no direct predecessors; ≥ 1, and an integer, otherwise;

$NDP(j)$ { = 0, if operation j has no direct predecessors; = j-$LPN(j)$, otherwise;

$NDS(j)$ { = -1, if $j = NO$ (indicating finished goods or another line segment); ≥ 2, and an integer, otherwise;

$A(j) > 0, (j = 1, 2, \ldots, NO)$;

$C(j) \geq 0, (j = 1, 2, \ldots, NO)$;

$U(j) \geq 1, (j = 1, 2, \ldots, NO)$;

$E(j) \geq 0, (j = 1, 2, \ldots, NO)$;

$N(j) \geq 0, (j = 1, 2, \ldots, NO)$.

The general production line configuration is characterized by a line in which serial and/or parallel segments occur. Therefore, in creating an optimizing model for this general configuration, it is useful to partition the operations j:

$L = \{j \mid NDP(j) = 0\}$

= set of operations which have no predecessors;

$M = \{j \mid NDP(j) \geq 1\}$

= set of operations which have predecessors.

Further, subsets I will be defined such that:

$I(j) = \{i \mid NDS(i) = j;\ j \epsilon M\}$

= set of immediate predecessors for operation $j, j \epsilon M$.

The complete statement of the optimizing model for the general configuration of the production line can be then written as shown below.

$$\text{Minimize} \quad Z = \sum_{j \epsilon L} A(j)*C(j)*N(j) \qquad \text{(a)}$$

$$+ \sum_{j \epsilon M} \left\{ C(j) - \sum_{i \epsilon I(j)} [C(i)*U(i)] \right\} *A(j)*N(j) \qquad \text{(b)}$$

$$+ \sum_{j=1}^{NO} C(j)*E(j); \qquad \text{(c)}$$

subject to:

$$\sum_{j=1}^{NO} N(j) = NW;$$

$$[U(j)*A(k)*N(k)] - [A(j)*N(j)] \leq E(j), (j = 1, 2, \ldots, NO-1).$$

where $k = NDS(j)$;

and $N(j) \geq 0, \quad (j = 1, 2, \ldots, NO)$.

The objective function is shown to consist of three parts:

(a) a variable term representing the total increase in the total cost of holding the in-process inventory produced during the time period on the operations having no direct predecessor(s);

(b) a variable term representing the net change in the total cost of holding the in-process inventory produced during the time period on the operations having direct predecessor(s); and

(c) a constant term representing the total cost of holding the in-process inventories present at the start of the time period.

Therefore it is shown that the production line, regardless of configuration, can be represented by an optimizing model made up of:

- an objective function that consist of, at most, NO terms plus a constant
- a manpower constraint of exactly NO terms;
- NO-1 flow constraints, each consisting of exactly two terms;
- the usual feasibility constraints.

This model can then be used to create fractional manpower assignments by using linear programming or integer manpower assignments by using integer programming.

Sample Problem

Consider the sample problem defined by the product tree shown in Fig. 1 and the data provided in Table 1. Partitioning the operations yields:

$L = \{1, 2, 3\}$ and

$M = \{4, 5, 6\}$.

The subsets I would then be:

$I(4) = \{1, 2\}$,

$I(5) = \{3\}$, and

$I(6) = \{4, 5\}$,

Table 1: Example problem.

j	1	2	3	4	5	6
$LPN(j)$	-1	-1	-1	1	3	4
$NDP(j)$	0	0	0	2	1	2
$NDS(j)$	4	4	5	6	6	-1
$A(j)$	40	50	60	30	50	20
$C(j)$	8	9	7	27	17	0
$U(j)$	1	1	1	2	1	1
$E(j)$	40	10	60	20	20	0

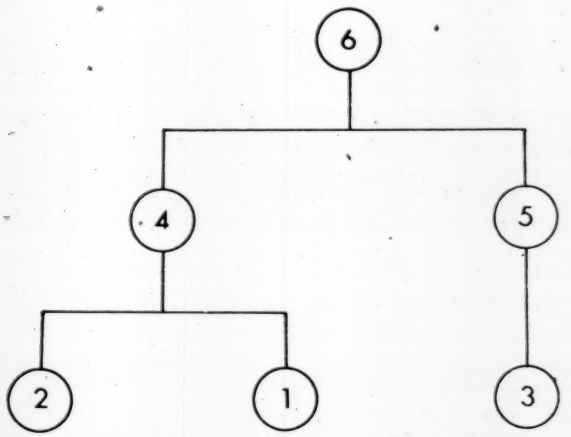


Fig. 1. Product Tree of Sample Problem.

Assuming that the number of workers to be assigned equals 24 and solving the problem using linear programming yields the solution vector shown below.

j	1	2	3	4	5	6
$N(j)$	4.36	4.08	0.62	7.14	1.94	5.86

These fractional assignments provide a Z value of 0.0 and assure zero in-process inventories at the end of the period with a total production of 117.2 units.

The solution to such a problem will not always provide zero in-process inventories at the end of the period. If the production line is out of balance such that one or more of the in-process inventories at the start of the period are large in comparison to the average usage at that operation, the resulting solution would assign zero manpower to the operation(s) in question. Even with such an assignment, the ending in-process inventory may be positive. For example, if we create a second sample problem by letting $E(3) = 150$, the solution vector becomes as shown below.

j	1	2	3	4	5	6
$N(j)$	4.49	4.19	0	7.33	2.00	5.99

This yields a Z value of 351.03, a total production of 119.8 units and in-process inventories zero, except at operation 3 which contains 50.15 units.

Using an integer program to solve the same two problems yields the solutions shown in Table 2. As expected, the Z values are higher for the integer solutions.

Table 2: Integer Programming solutions.

	Problem 1		Problem 2	
j	$N(j)$	$E(j)$	$N(j)$	$E(j)$
1	5	30	6	70
2	4	0	4	0
3	1	20	0	50
4	7	30	7	30
5	2	20	2	20
6	5	100	5	100
	$Z = 1530$		$Z = 2060$	

It should be noted that the linear programming technique can not be used directly as it results in fractional assignments which theoretically maintain all in-process inventories at a zero level. Since the production rate of each operation is different, it would be impossible to implement such a solution unless buffers were used at each in-process inventory. Even with the use of buffers it may not be obvious as to when and where a worker is to be moved. This does not mean that a worker must remain at the same operation for an entire shift. The production schedule could use integer programming and a two hour time period which would result in four separate schedules for an entire shift. Thus a worker could change operations during a shift, but only if an appropriate time period was used in the development of the shift schedule. Otherwise a possible shortage could occur at one or more of the operations.

Although the linear and integer programming techniques do provide the optimal solutions, they may be impractical for implementation. In practice, the problem would have to be resolved at the start of each new scheduling period. In general, the problem will be the same with only the in-process inventories and number of workers changing. The in-process inventories can be estimated using the solution from the previous period. Unfortunately, the exact number of available workers may be a variable and may not be known until just before the start of the scheduling period. Therefore a computer and associated code would have to be readily available at the start of each scheduling period. An alternative would be to solve several problems by varying the number of workers over the estimated number which are anticipated to be available. It is also possible that a company, particularly a small company, will not have a computer, a linear programming code, or an integer programming code readily available. Consequently, it would be desirable to have a technique which would not require extensive computer time and if possible, require only pencil and paper with the possible inclusion of a hand calculator.

The following algorithms are provided as an alternative to the linear integer programming techniques. One may

question the usefulness of an alternative fractional assignment procedure, since it is not possible to make fractional assignments in a real world application. However, the fractional assignment algorithm does provide an alternative to linear programming for this special case and this initial solution is subsequently used as a starting point for the integerizing algorithm.

The Fractional Assignment Algorithm

The three vectors, *LPN*, *NDP*, and *NDS*, completely define a production line configuration. They were used in the previous formulation because of convenience in developing the optimizing model. Another method of completely defining the configuration of a production line is that of a total utilization matrix, T, where:

$t(j, k)$ = total number of part j required in part k.

Other words, to produce one final product, requires $t(j, NO)$ of part j, etc., for all j.

In order to maintain the in-process inventories at their existing level, each worker assigned to the final operation, NO, requires

$$X(j) = t(j, NO) * A(NO)/A(j), \quad (j = 1, 2, \ldots, NO\text{-}1)$$

workers assigned to the remaining, NO-1, operations. A general expression for the number of workers assigned to the first NO-1 operations as a function of the number of workers assigned to the last operation, NO, can then be written as follows:

$$X(j) = X(NO) * A(NO) * \left[\frac{t(j, NO)}{A(j)}\right], \quad (j=1,2,\ldots,NO).$$

In order to assign exactly NW workers to the line the following expression must hold:

$$\sum_{k=1}^{NO} X(k) = X(NO) * A(NO) * \sum_{k=1}^{NO} \frac{t(k, NO)}{A(k)} = NW .$$

Solving this expression for $X(NO)$ and substituting into the previous expression yields:

$$X(j) = [NW * t(j, NO)] / \left[A(j) * \sum_{k=1}^{NO} \frac{t(k, NO)}{A(k)}\right], \quad (j=1,2,\ldots,NO) .$$

This provides an expression for determining the fractional manpower assignment which will maintain the existing in-process inventory levels.

If one desires to obtain the assignment which will yield zero-in-process inventories, or as close as possible, one must consider the levels of the existing inventories. The equivalent number of part j represented by all in-process inventories, $SET(j)$, can be computed as follows:

$$SET(j) = \sum_{k=j}^{NO} E(k) * t(j, k) \qquad (j=1,2,\ldots,NO).$$

Using these values, it is then possible to reduce, by the appropriate value, the assignments at the stations which have a positive equivalent inventory as follows:

$$XM(j) = X(j) - \frac{SET(j)}{A(j)}, \qquad (j=1,2,\ldots,NO) .$$

Where $XM(j)$ represents the minimum manpower allocation to achieve zero in-process inventory levels.

If all the starting in-process inventories had been zero, the sum of the new assignments would be equal to NW. However, this would normally not be the case; thus, one must now readjust the assignments to reflect the reductions by the existing in-process inventories. This is accomplished by computing the total reduction, DX, as follows,

$$DX = NW - \sum_{j=1}^{NO} XM(j) ,$$

and then normalizing this allocation, DX, over the entire line. This is accomplished by using the following expression,

$$XA(j) = XM(j) + DX * X(j)/NW, \qquad (j=1,2,\ldots,NO) .$$

where $XA(j)$ is the fractional manpower allocation to operation j that would result in all in-process inventories becoming zero, or as close as possible to zero, at the end of the period.

The steps of the formalized algorithm are as follows:

1. Construct the total utilization matrix T.
2. Compute $STOA = \sum_{j=1}^{NO} \frac{t(j, NO)}{A(j)}$
3. Compute $X(j) = \frac{NW * t(j, NO)}{STOA * A(j)}, \quad (j=1,2,\ldots,NO)$.
4. Compute $SET(j) = \sum_{k=j}^{NO} E(k) * t(j, k), \quad (j=1,2,\ldots,NO)$.
5. Compute $XM(j) = X(j) - \frac{SET(j)}{A(j)}, \quad (j=1,2,\ldots,NO)$.
6. Compute $DX = NW - \sum_{j=1}^{NO} XM(j)$

7. Compute $XA(j) = XM(j) + (DX * X(j)/NW)$,

$(j = 1, 2, \ldots, NO)$.

Application of this algorithm requires only a single pass or a known number of steps. In contrast, linear programming is an iterative technique which requires a generally unknown number of steps. In addition, the proposed algorithm is far easier to use if only pencil and paper or a small calculator are available.

In using the algorithm to solve the original sample problem, the first step is to construct the total utilization matrix, T, which is given below:

$$T = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 2 \\ 0 & 1 & 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The remaining steps of the algorithm are summarized in Table 3. The resulting solution is the same as the solution obtained by linear programming.

Table 3: Summary of Solution Values.

J	$X(j)$	$SET(j)$	$XM(j)$	$XA(j)$
1	4.932	60	3.432	4.36
2	3.945	30	3.345	4.08
3	1.644	80	0.310	0.62
4	6.575	20	5.908	7.14
5	1.973	20	1.573	1.94
6	4.932	0	4.932	5.86

If we consider the situation where we let $E(3) = 150$ and resolve the problem the solution vector becomes as shown below.

j	1	2	3	4	5	6
$XA(j)$	4.52	4.21	-0.12	7.36	2.01	6.02

It should be noted that a negative assignment exists and that although the sum of the assignments equals 24, the sum of the positive assignments equals 24.12. Other words, the algorithm, as stated, will not always yield the same solution as one would obtain from linear programming. The equivalent solution can be obtained by setting all negative assignments to zero and then normalizing the necessary reduction over the remaining positive values.

In actual practice this would probably not be necessary since it is not possible to make fractional manpower assignments. Therefore it is necessary to convert the feasible fractional assignment to a feasible integer assignment; while at the same time maintaining a minimum or near minimum in-process inventory cost. For most problems, particularly large problems, this can be a very difficult task since the inventories are dependent on the assignments at both the succeeding as well as the preceding operations. One alternative is to use the initial formulation and employ an integer programming code. For those instances where integer programming is not an acceptable alternative, the following algorithm is provided.

The Integer Assignment Algorithm

The algorithm assumes that the optimal fractional manpower allocations, $XA(j)$'s, from the previous algorithm are available as a starting point. This fractional solution is then converted to a feasible integer solution by the interchange of workers with the next operation each time an infeasible allocation is found.

The actual procedure starts with the fractional assignments and truncates them to obtain starting integer values. If negative assignments exist, they are set equal to zero. This will normally result in the number of workers assigned being less than required. Let the number of workers not assigned equal NS. These are arbitrarily assigned to the first NS operations, one per operation. The reader should note that it is always possible to add workers to operations which have no predecessors.

Starting with the first operation, the assignment is checked for feasibility. If the assignment is found to be infeasible, then a worker is removed from that operation's direct successor and added to the assignment at the operation in question. Each time a change in assignments is made, all previously made assignments must be rechecked for feasibility. If an assignment is found to be feasible, one proceeds to the next operation. This process is continued until a feasible integer solution is created. The resulting integer solution is not necessarily optimal, although initial experiments have shown the solutions to be quite good.

The actual algorithm can be stated as follows:

1. Set $N(j) = \begin{cases} \text{greatest integer in } XA(j) \text{ if } XA(j) \geq 0 \\ 0, \text{ otherwise} \end{cases}$

$(j = 1, 2, \ldots, NO)$

2. Compute $NS = NW - \sum_{j=1}^{NO} N(j)$

3. Set $N(j) = N(j) + 1$, $(j = 1, 2, \ldots, NS)$

4. Set $j = 1$

5. Set $k = NDS(j)$

6. Compute $E(j) = E(j) + A(j) * N(j) - U(j) * A(k) * N(k)$

If $E(j) < 0.$, go to step 7

If $E(j) \geq 0.$, go to step 8

7. Set $N(j) = N(j) + 1$
   Set $N(k) = N(k) - 1$
   Go to step 4

8. Set $j = j + 1$

   If $j \leq NO - 1$, go to step 5
   If $j > NO - 1$, go to step 9

9. End .

Table 4: Solutions from Integer Algorithm.

	Problem 1		Problem 2	
j	$N(j)$	$E(j)$	$N(j)$	$E(j)$
1	5	30	5	30
2	5	50	5	50
3	1	20	0	50
4	7	70	7	30
5	2	40	2	20
6	4	80	5	100
	$Z = 3400$		$Z = 2190$	

Using this algorithm to solve the two example problems yields the solutions shown in Table 4. Contrasting these solutions to those obtained from the integer programming technique reveals that they are quite similar. The objective value obtained for the first problem by the Integer Algorithm is greater than twice the value from that obtained from the integer program. This is due to the fact that a large decrease in the objective value is attained when additional workers are assigned to the last operation, as the incremental cost is negative. The primary difference between the two solutions occurs at station 2 and station 6. This accounts for the discrepancy in the Z values. The solutions to the second problem are very similar and differ only in the assignments made in the first two stations.

Concluding Remarks

It is rather obvious that the solution to a one period problem obtained from the integer algorithm can only be as good and never better than that obtained from integer programming. Although the integer programming code does provide the optimal solution for a single period it will not necessarily provide the optimal solution for a multiple period time frame which is a more accurate measure of cost associated with the in-process inventories on a production line. It would be possible to formulate a multiple period integer program; unfortunately the size of the problem quickly becomes rather prohibitive and one would have to estimate the number of workers available for each period. Initial tests have shown that for multiple problems the integer algorithm provides total cost solutions that are very similar, and sometimes better than those achieved by using integer programming. One must also consider the time required to obtain a solution. For medium to large size production lines that contain in excess of fifteen stations the time required to obtain the integer programming solution can easily be in excess of a hundred times the computational time required to achieve the solution from the proposed algorithm.

In general it is not possible to make direct time comparisons between the two techniques because the time required to solve a given problem using integer programming is not only dependent upon the size of the problem, but also the structure, order of variables, order of constraints and the algorithm code. However, such time comparisons become relatively unimportant when one considers that the integer algorithm can be easily applied without the use of a computer. Once the user becomes familiar with the steps of the algorithm it is possible to develop a schedule for a line with ten to fifteen operations in less than ten minutes using only pencil and paper. Use of a calculator would further reduce this time. Therefore, one of the principle attributes of the integer algorithm is its ease of application.

In the actual implementation of this technique one should undoubtedly consider using buffer inventories during the initial stages to avoid the possibility of running out of material at a particular station. This could conceivably happen due to the fact that it is assumed that production rates at each station are deterministic; whereas in a real life situation they are more likely to be random variable. However, it should be possible to determine reasonable buffers to account for this randomness in the production rates fairly quickly. In addition, the algorithm requires a particular system of numerical designation of the operations. This designation requires that the first operation be labelled as number 1 and that the last operation carries the highest number designation. In addition, consecutive integers should be used to designate the set of operations having a common direct successor. For example, in the sample problem operations 1 and 2 have a common successor, operation 4. As required, these operations are consecutive integers. The same designation pertains to operations 4 and 5 which have a common successor, operation 6. In addition the operations at any given level in the product tree are designated as a set of consecutive integers. It has also been observed that better solutions are obtained when the operation giving the lowest designation in any set has the lowest incremental cost. For example, operation 4 has an incremental cost of 10 which is the same as the incremental cost of operation 5. Consequently the numbering of these two operations could be reversed without effecting the solution obtained from the algorithm. The reader should note that in the event that any of the costs changes it may be necessary to renumber the product tree.

In actual practice it may be very difficult to obtain the exact cost of the various in-process inventories. However, in using this algorithm it is not necessary to have exact costs. It is only necessary that the costs are relative to each other. For example, in the sample problem the same solution would have been obtained had all of the costs been multiplied or divided by a constant value. Consequently, in actual

practice it may be fairly easy to come up with acceptable cost values by taking into consideration some ratio of the space required to store the product and the actual cost of the product at that point in the production sequence.

In order to obtain the integer assignment at the start of a particular time period it is necessary to compute the initial fractional manpower assignments by using the fractional assignment algorithm and then using these solutions as input to the integer assignment algorithm solve for the integer assignments. It should be noted that it is only necessary to use the first three steps of the fractional assignment algorithm once, as these values will remain constant as long as the product tree, cost values, and number of workers remain constant. Consequently it is possible to start at step 4 in the fractional manpower assignments as long as the only values changing are the in-process inventories and on step 3 if the number of workers also varies.

It is the opinion of the authors that the proposed algorithms provide a viable alternative to linear and integer programming techniques for the unbalanced production line problem. This is particularly true for a company which does not have the access to either a computer or associated codes.

Acknowledgment

The authors would like to thank Dick McGee of Digital Equipment Corporation for his assistance and support of this research.

References

[1] Anderson, David R., "Simulation of Sequential Production Systems with In-Process Inventory," Winter Simulation Conference of ACM, 6th Annual, San Francisco, pp. 85-92 (January 1973).

[2] Davis, K. Roscoe, and Taylor, Bernard W., "Want to Reduce WIP Inventory? Try Linear Line-Balancing," *Production and Inventory Management*, Second Quarter, pp. 57-65 (1974).

[3] Davis, K. Roscoe, "Interactive Simulation: A Solution to the Inventory Problem," *Industrial Engineering*, **7** 6, pp. 30-35 (June 1975).

[4] Jensen, P. A., and Khan, H. A., "Scheduling in a Multistage Production System with Set-up and Inventory Costs," *AIIE Transactions*, **4** 2, pp. 126-133 (June 1972).

[5] Kraemer, Svend A., and Love, Robert F., "A Model for Optimizing the Buffer Inventory Storage Size in a Sequential Production System," *AIIE Transactions*, **II** 1, pp. 64-69 (March 1970).

[6] Montgomery, Douglas C., and Berry, William L., Editors, *Production Planning, Scheduling, and Inventory Control: Concepts, Techniques, and Systems*, American Institute of Industrial Engineers (1974).

Dr. Russell E. Jacobson is an Associate Professor in the Management Department at the University of Wisconsin–Whitewater. He holds a BSEE from the University of Wisconsin–Madison, a MS in Management from Rensselaer Polytechnic Institute, and a PhD in IE/OR from the University of Massachusetts. His research interests are in the area of production planning and control. He is a member of AIIE, APICS, Alpha Pi Mu, and Tau Beta Pi.

Dr. Randall P. Sadowski is an Assistant Professor in the School of Industrial Engineering at Purdue University. Previously he was with the Department of Industrial Engineering and Operations Research at the University of Massachusetts. He holds a BSME and MSIE from Ohio University and a PhD from Purdue. His research interests are in the areas of production and manufacturing and systems, and computer applications in these areas. He is a Senior Member of AIIE and also a member of APICS.

A Multi-Resource Project Scheduling Algorithm

E. M. DAR-EL
SENIOR MEMBER, AIIE

YIGAL TUR
Technion-Israel Institute of Technology
Haifa, Israel

Abstract: A heuristic algorithm for scheduling a multi-resource project is described where the objective is to minimize the project cost for any project duration between some specified time range. Input conditions are realistic and include alternative resource combinations (ARCs) for each activity, resource cost functions and hiring and dismissal costs for changes in resource quantities. The algorithm selects that combination of activity ARCs that gives the minimum cost solution for any feasible project duration, and eventually generates the Variable Cost-Time trade-off relationship from which the minimum cost conditions for the project can be found.

■ The problem of resource allocation in real projects becomes more complicated as additional resource types are involved. This is because of the combinatorial nature of the problem, where a good solution must be selected from a very large number of feasible solutions.

The possibility of using a mathematical model to solve this problem in real life depends both on its applicability as well as on obtaining optimal or good solution economically.

Research works dealing with project resource allocation may be categorized as follows: Resource levelling problems, which aim at obtaining a smooth resource profile for a given project duration; and problems which aim at minimizing project time or project cost without violating fixed upper bounds for resource quantities.

Various methods have been developed to tackle these types of problems:

Zero-one programming methods include those reported in [6], [8], [16], [17] and [21], but solutions are difficult to obtain for real-life projects because of computational limitations. Branch and bound techniques are used in [2], [7], [10] and [18], and a dynamic programming approach was proposed by Petrovic [15], who used a 'weighting factor' to select between resource types in a multi-resource project. Heuristic models are described in [11], [20], [21] and [22]. Of particular interest is Wiest [21], who introduced the idea of scheduling on a daily basis. Hooper [9] and Pascoe [13] distinguish between 'serial' and 'parallel' methods in allocating resources. In the serial ones, activities are selected independently, whereas in parallel methods, activity combinations are selected. Parallel methods are more efficient, but require more computation time. Three important surveys on constrained resource project scheduling have recently appeared. These are, two articles by Davis [3], [4] and one by Davis and Petterson [5].

Actually, none of the above works adequately reflect the conditions that occur in real projects, especially in relation to the following factors:

(a) *Alternative Resource Combinations (ARC's).* An activity may require a combination of different resources for its performance. It may also happen that the same activity can be executed by using any of several such combinations, each one having its own completion time. For example, a construction activity can be performed by 2 riggers plus 2 helpers using block and tackle, taking 6 hours for its performance. It may also be done in 1 hour by employing a crane together with 1 rigger plus 1 helper. These combinations are referred to as 'Alternative Resource Combinations,' or ARCs.

(b) *Resource Cost Functions.* It is generally assumed

Received July 1975; revised February 1976; July 1976.

that a linear cost-quantity relationship exists for each resource. However, the scarcity of some resource types may result in a concave, nonlinear or piecewise linear relationship, as illustrated in Figure 2.

Such nonlinear relationships are implied in some forms of activity cost-time trade-off data, but the selection of each resource quantity occurs independently, whereas in practice it is the combination of resource types (i.e. ARCs) and their accumulated individual totals that must be considered.

(c) *Hiring and Dismissal of Resources.* The required total quantity of each resource usually fluctuates during the execution of a project. These fluctuations are always associated with some additional costs in the form of 'hiring' or 'dismissing' the resources. Such costs not only apply to projects where resources are subcontracted, but also to companies that own their resources, the difference being reflected in the magnitude of the hiring and dismissal costs.

A zero-one IP formulation was constructed [19] which takes these three factors into account. It was shown that the number of variables and constraints can be very large, growing exponentially with the number of activities. For example, a 50-activity project having 3 resource types, an average of 3 ARCs per activity and a project duration of about 50 days, requires some 3200 binary variables and 1500 constraints which makes the solution method unfeasible.

Consequently, the aim of this study is to develop a heuristic algorithm for finding the best conditions for the project execution when the input data is used in its most practical form, viz.: as several ARC's for each activity; compilations of resources cost functions; and estimation of hiring and dismissal costs for each resource.

1. Model Description

Consider a network consisting of n activities connected by precedence constraints and represented by an 'Activity-On-Node' diagram, as illustrated in Fig. 1.

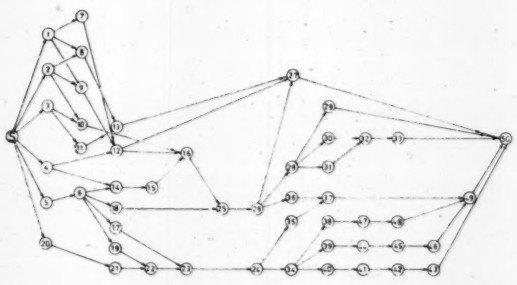


Fig. 1. 50-activity AON project network.

1.1 Each activity i ($i = 1, \ldots, n$) can be completed by any one of a specified number of different alternative resource combinations, or ARC's. Each ARC has p+1 parameters, listed as p elements (i.e. quantity of each resource type) of a three dimensional resource matrix (called the SORCE matrix), and one parameter representing the time duration. The (ijk)*th* element of the SORCE matrix is the quantity of resource k ($k = 1, \ldots, p$) needed for the completion of activity i (i=1, . . . , n) by using ARC j ($j = 1, \ldots, L_i$, where L_i is the number of ARC's associated with activity i). When $k = p$+1, the element (ijk) gives the duration for completing activity i using ARC j. An illustration of the SORCE matrix is shown in Table 1.

Table 1: A sample of the SORCE matrix k=1, 2, . . . , p, and p=3.

Activity number i	Alternative ARC j	Resource Amounts k=1	k=2	k=3	Duration k=p+1
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
17	1	5	10	4	5
	2	4	15	3	4
	3	3	18	3	4
	4	2	20	2	3
18	1	4	20	2	8
19	1	3	20	0	6
	2	3	10	2	4
.	.	.	.	.	.
.	.	.	.	.	.

The Basic Schedule Set (or simply Basic Set) is defined as a set of n ARC's, which uniquely specifies the intended method of executing the project. Considering the data in Table 1 for example, one Basic Set may include the following ARCs:

activity	ARC number
.	.
.	.
.	.
17	4
18	1
19	2
.	.
.	.
.	.

The number of Basic Sets possible for any practical project can be very large indeed.

1.2 Changes in resource levels with respect to time are associated with hiring and dismissal costs.

The hiring cost for resource k consists of two parts: a fixed cost g_{1k}, and a variable cost proportional to the resource level difference, i.e. g_{2k} ($q_{kd} - q_{k,d-1}$), where q_{kd} is the quantity of resource k used on day d. Thus, hiring

cost associated with resource k on day d is:

$$\left.\begin{array}{ll} g_{1k} + g_{2k}(q_{kd} - q_{k,d-1}) \ldots & \text{if } q_{kd} > q_{k,d-1} \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots & \text{if } q_{kd} \leq q_{k,d-1} \end{array}\right\} \quad (1)$$

The dismissal cost for resource k consists of two parts: a fixed cost h_{1k}, and a variable cost proportional to resource level difference, i.e. $h_{2k}(q_{k,d-1} - q_{kd})$. Thus, dismissal cost associated with resource k on day d is:

$$\left.\begin{array}{ll} h_{1k} + h_{2k}(q_{k,d-1} - q_{kd}) \ldots & \text{if } q_{kd} < q_{k,d-1} \\ 0 \ldots\ldots\ldots\ldots\ldots\ldots & \text{if } q_{kd} \geq q_{k,d-1} \end{array}\right\} \quad (2)$$

1.3 The cost function for the k*th* resource is assumed to be a piecewise linear curve (having M_k linear sections), as shown in Fig. 2.

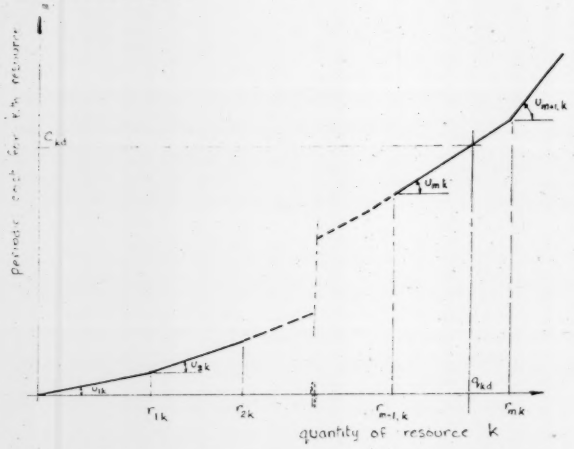


Fig. 2. Illustrating a typical piece-wise linear cost function.

Then C_{kd}, the total cost of resource k on day d is calculated as follows:

Consider resource k needed in quantities between $r_{m-1,k}$ and $r_{m,k}$ and having unit cost u_{mk}. Then

$$C_{kd} = q_{kd}(u_{mk} + g_{2k}x_{1k} - h_{2k}x_{2k}) + \sum_{i=1}^{m-1}(u_{ik} - u_{i+1,k})r_{ik}$$

$$+ g_{1k}x_{1k} + h_{1k}x_{2k} + q_{k,d-1}(h_{2k}x_{2k} - g_{2k}x_{1k}) \quad (3)$$

$$\text{for } r_{m-1,k} \leq q_{kd} < r_{mk}\,.$$

Where u_{mk} is the unit cost of resource k operating in the m*th* step of the resource cost curve.

r_{mk} is the upper bound of the resource quantity in the m*th* region.

x_{1k}, x_{2k} are binary variables, defined as follows:

$$\left.\begin{array}{l} x_{1k} = \begin{cases} 1 \text{ ------- if } q_{kd} > q_{k,d-1} \\ 0 \text{ ------- if } q_{kd} \leq q_{k,d-1} \end{cases} \\ x_{2k} = \begin{cases} 1 \text{ ------- if } q_{kd} < q_{k,d-1} \\ 0 \text{ ------- if } q_{kd} \geq q_{k,d-1} \end{cases} \end{array}\right\} \quad (4)$$

$$\text{for } k = 1, \ldots, p.$$

To illustrate these costs, let the Hiring, Dismissal and Resource cost data for the 3 resources indicated in Table 1, have the values given in Tables 2 and 3. Figure 3 also illustrates the cost function for resource $k = 3$ (data from Table 3).

Table 2: Hiring and dismissal costs.

k	g_{1k}	g_{2k}	h_{1k}	h_{2k}
1	25	25	15	20
2	10	20	5	10
3	50	100	25	50

Table 3: Resource cost data.

$k = 1$				$k = 2$	
r_{11}	u_{11}	r_{21}	u_{21}	r_{12}	u_{12}
10	100	00	400	00	75

$k = 3$							
r_{13}	u_{13}	r_{23}	u_{23}	r_{33}	u_{33}	r_{43}	u_{43}
2	100	5	200	10	400	00	800

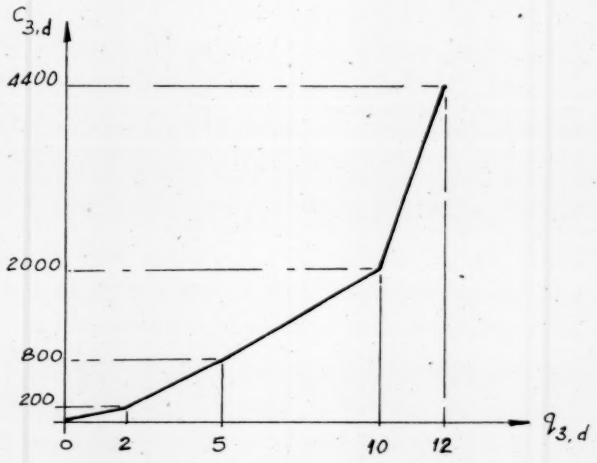


Fig. 3. Cost function for resource type k=3.

Assume that from day 4 to day 5, the quantity needed of resource k=3 is increased from 6 units to 9 units.

Then from Eq. (3), the total cost for resource 3 on day 5 is $C_{3,5}$ and is given by:

$$C_{3,5} = 9(400 + 100.1 - 50.0) + (100-200)\,0.2 + (200-400)5 + 50.1 + 25.0 + 6(50.0-100.1)$$

$$= 2750$$

1.4 A 'weighting factor' vector μ ($\Sigma_{k=1}^{p}\ \mu_k = 1$) is defined. μ_k represents the relative weight (i.e. the importance), given to the k*th* resource, thus providing a criterion for resolving conflicting demands that may arise between the different resource types. There are many ways in which the weighting factors can be determined. The following method, based on relative resource marginal costs, was used in the algorithm for calculating μ_k:

$$\mu_k = \frac{\text{maximal slope of } k\text{-resource cost function}}{\text{sum of } p \text{ maximal slopes}} \qquad (5)$$

$$\text{where } k = 1, \ldots, p \text{ and } \sum_{k=1}^{p} \mu_k = 1.$$

This method of calculating μ_k was selected in order to avoid selecting activity combinations that require resources having high marginal costs.

1.5 Activity splitting is permitted, but these incur some 'penalty.' The magnitude of this penalty can be such, as to avoid splitting altogether. Splitting penalties are further discussed in Section 6.

1.6 Critical activities can be delayed, but penalties may be imposed, whose magnitude is such as to either permit delays, or else avoid them altogether. Delay penalties are discussed further in Section 6.

2. The Objective Function

The aim of this study is to develop a heuristic algorithm, which finds the basic set which minimizes the total variable costs, for every duration over a specified time range.

The number of basic sets that can be generated is usually very large, and therefore an efficient method must be devised for checking a small part of this number.

The final outcome of this program is to generate the traditional 'project variable cost–project duration' relationship, which, when combined with the fixed costs, will indicate the optimal project duration. (Figure 4 illustrates the 'project variable cost–project duration' trade-off relationship for the entire range of project durations possible, for the 50-activity network in Fig. 1 using the data (shown in part) given in Tables 1, 2, and 3. The curves were generated for the conditions of zero and infinite delay penalties (for critical activities) and for zero and infinite activity splitting penalties. The total range of project durations is from 52 to 98 days, i.e. a range of 46 days.

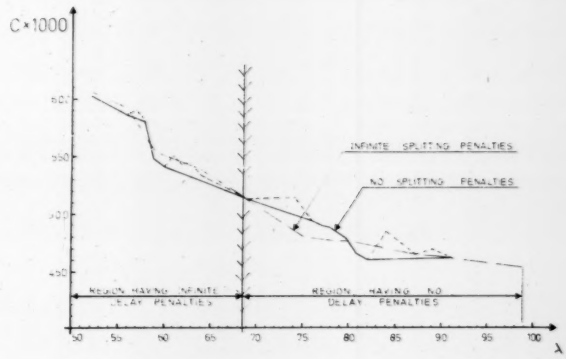


Fig. 4. Project cost curve.

It is noted that for some project durations, the infinity-splitting penalty solutions are better than those for the no-splitting penalty case. This is due to the combinatorial nature of the problem, where it may happen that the infinity-splitting rule forces an early (but poorer assignment than for the no-splitting case) allocation of an activity, but which eventually leads to a more efficient allocation for most of the remaining work days.

3. Algorithm Input Data

Input data for the algorithm include the following:

3.1 Network data – the set of immediate followers for each activity.

3.2 Activity data – the set of ARC's for each activity, each ARC specifies the amount of each resource and the ARC duration.

3.3 Hiring and dismissal costs for each resource type.

3.4 Cost functions for each resource type.

3.5 Splitting penalty coefficient for each resource type.

3.6 Critical activity delay penalty.

Figure 5 illustrates the overall process flow chart and the algorithm input data is identified by the box marked ①. Box ④ to ⑨ represent the generation of Basic Sets and is discussed in Section 5. As each Basic Set is created, the algorithm proceeds to the resource allocation process, where activities are scheduled on a day-to-day basis. This process is discussed in section 6.

4. Initial Calculations

The following initial conditions occur in box marked ② in Fig. 5.

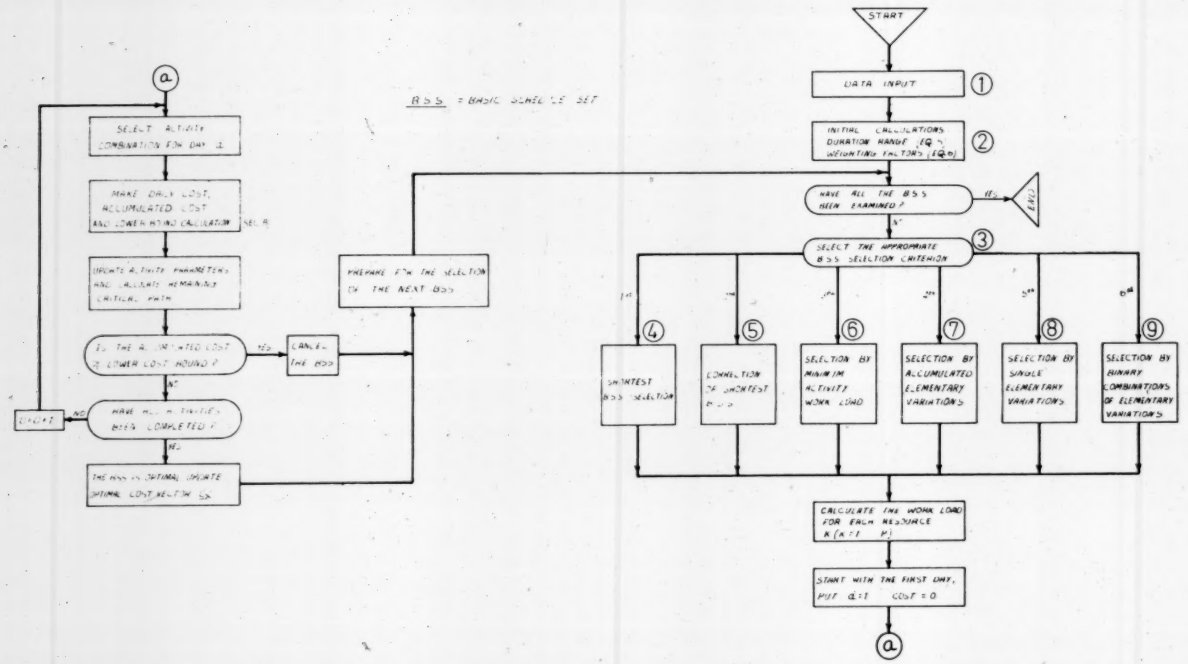

START
DATA INPUT
INITIAL CALCULATIONS
DURATION RANGE (EQ.)
WEIGHTING FACTORS (EQ.)
HAVE ALL THE BSS BEEN EXAMINED ?
SELECT THE APPROPRIATE BSS SELECTION CRITERION
SHORTEST BSS SELECTION
CONNECTION OF SHORTEST
SELECTION BY MINIMUM ACTIVITY WORK LOAD
SELECTION BY ACCUMULATED ELEMENTARY VARIATIONS
SELECTION BY SINGLE ELEMENTARY VARIATIONS
SELECTION BY BINARY COMBINATIONS OF ELEMENTARY VARIATIONS
CALCULATE THE WORK LOAD FOR EACH RESOURCE
START WITH THE FIRST DAY, PUT COST = 0
BSS = BASIC SCHEDULE SET
SELECT ACTIVITY COMBINATION FOR DAY
MAKE DAILY COST, ACCUMULATED COST AND LOWER BOUND CALCULATION
UPDATE ACTIVITY PARAMETERS AND CALCULATE REMAINING CRITICAL PATH
IS THE ACCUMULATED COST > LOWER COST BOUND ?
CANCEL THE BSS
HAVE ALL ACTIVITIES BEEN COMPLETED ?
THE BSS IS OPTIMAL UPDATE OPTIMAL COST, SELECTION ETC.
PREPARE FOR THE SELECTION OF THE NEXT BSS

Fig. 5. Overall Process Flow Chart.

4.1 The project duration range: In order to determine the duration range, two critical path lengths are calculated. The first uses the shortest duration ARC for each activity, which gives the lower bound λ_1 for the project duration λ while the second uses the longest duration ARC for each activity, and gives the upper bound λ_2 of critical path length.

Let $\bar{\lambda}$ be the width of the project duration range, then

$$\bar{\lambda} = \lambda_2 - \lambda_1 + 1 \qquad (6)$$

4.2 Weighting factors: These are calculated according to Eq. (5).

5. Basic Schedule Sets (Basic Sets)

The space of basic sets is very large, and only a small part of it can be examined. Efficient heuristic rules must therefore be applied in order to identify the basic set which will give the 'best' schedule.

The method of varying the basic sets determines the algorithm efficiency, since it controls the rapidity of approaching the 'best' solution and its associated computation time.

The process of basic sets advances as follows:

First define w_i, the weighted work load for each ARC included in activity i:

$$\text{i.e. } w_i = \sum_{k=1}^{p} \mu_k \text{ *SORCE}(i,j,p+1)\text{*SORCE}(i,j,k) \qquad (7)$$
$$(\text{for } i=1,\ldots,n)$$

where:

SORCE($i,j,p+1$) is the activity duration, if ARC j is selected,

SORCE(i,j,k) is the k*th* resource amount associated with activity i and ARC j.

The value of w_i indicates the 'equivalent quantity' of work contained in each ARC and is used as a selection criterion when choosing between ARCs.

5.1 Start with selecting the shortest ARC for each activity, i.e. without taking the resource demands into account. This gives the shortest possible project duration (i.e. λ_1) see box ④ in Fig. 5).

5.2 A trial is then made to decrease the total cost *without increasing* the project duration. Thus for each noncritical activity, determine that ARC having the lowest w_i value (Eq. 7) and which does not raise the minimum critical path length, λ_1. The new basic set therefore includes all original critical activity ARC's, plus all ARC's on noncritical activities, found according to the above criterion. (See box ⑤ in Fig. 5.)

5.3 A new criterion is now used for selecting the next basic set. For each activity i, select $(w_i)_{\min}$, the ARC having the minimal value of w_i (i.e. ignoring the activity duration). Thus the third basic set to be tested contains all ARC's yielding the minimum equivalent work loads. (See box ⑥ in Fig. 5.

5.4 Variations of the next group of basic sets to be tested are generated as follows: Consider any basic set. Let Δw_i be the smallest difference between $(w_i)_{\min}$ and the w_i value obtained from some other ARC in activity i. Find the ARC having the smallest Δw_i value, taken over all activities, and substitute it with the appropriate ARC (i.e. from the same activity) in the existing basic set. This change, called an elementary variation, produces a new basic set.

The number of elementary variations permitted is pre-defined, and its value clearly controls both the solution accuracy and computation time.

Starting with the third basic set defined in Section 5.3, elementary variations are introduced by using three different methods:

5:4.1 Elementary variations are introduced singly, meaning that when a new variation is created, the last one is canceled. (Box ⑧ in Fig. 5.)

5.4.2 Elementary variations are introduced accumulatively. (Box ⑦ in Fig. 5.)

5.4.3 By binary combinations of all such elementary variations up to some pre-defined limit. (Box ⑨ in Fig. 5.)

Rules 5.4.1 to 5.4.3 are illustrated in the following example.

5.5 *An example illustrating the changing of Basic Sets.*

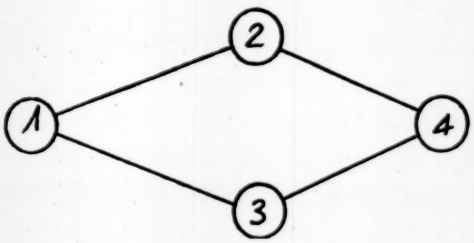
Consider the following 4-activity network with its accompanying SORCE input data.

Given:

$\mu_1 = 0.2$

$\mu_2 = 0.3$

$\mu_3 = 0.5$



SORCE (Input)								
Activity (i)	ARC (j)	Resource demand k=1	k=2	k=3	Duration (p+1)=4	w_i	Δw_i and (Rank)	
1	1	4	8	1	9	33.3	1.1	(1)
	2	3	12	0	8	33.6	1.4	(2)
	3	3	10	2	7	32.2	-	
2	1	3	12	1	8	37.6	3.3	(3)
	2	2	10	3	7	34.3	-	
	3	4	12	4	6	38.4	4.1	(5)
3	1	3	8	2	12	48.0	-	
4	1	4	8	0	8	25.6	-	
	2	2	10	3	6	29.4	3.8	(4)

From Eq. (3):

$$w_i = \{0.2\,\mathrm{SORCE}(i,j,1)+0.3\,\mathrm{SORCE}(i,j,2)+0.5\,\mathrm{SORCE}(i,j,3)\} \times \mathrm{SORCE}(i,j,p+1)\,.$$

For example, Δw_1 for ARC No. 1 is 33.3-32.2 = 1.1. It is ranked as (1) because its w_i value is the least for the entire set of activities.

(a) The first Basic Set (from 5.1) gives the shortest duration

$\lambda = \lambda_1 = 25$

BASIC SET 1	
Activity	ARC
1	3
2	3
3	1
4	2

(b) The second Basic Set maintains $\lambda_1 = 25$, but non-critical activity 2 has its ARC No. 3 replaced by ARC No. 2, since its w_i is less and λ_1 is unchanged (see 5.2).

BASIC SET 2	
Activity	ARC
1	3
2	2
3	1
4	2

(c) The third Basic Set includes those ARCs having the least w_i values (see 5.3).

$\lambda = 27$

BASIC SET 3	
Activity	ARC
1	3
2	2
3	1
4	1

(d) Starting point for these generations is Basic Set 3. Basic Sets 4 to 8 occur through *single* replacements of ARC's starting from the least value of Δw_i (see 5.4.1). Three ARCs from Basic Set 3 must always be present in every one of these new Basic Sets. NOTE: Basic Set 7 is repeat of Basic Set 2, and is therefore not utilized.

BASIC SET NO.	4	5	6	7	8
Activity	ARC	ARC	ARC	ARC	ARC
1	•1	•2	3	3	3
2	2	2	•1	2	•3
3	1	1	1	1	1
4	1	1	1	•2	1

• Changes

(e) Starting point for these generations is Basic Set 3. Basic Sets 9 to 13 occur through accumulative replacement of possible ARCs for each activity (5.4.2).

Basic Sets 9 and 10 are repeats of Basic Sets 4 and 5 and are not utilized.

BASIC SET NO.	9	10	11	12	13
Activity	ARC	ARC	ARC	ARC	ARC
1	•1	•2	•2	•2	•2
2	2	2	•1	•1	•3
3	1	1	1	1	1
4	1	1	1	•2	•2

• Changes

(f) Basic Sets 14 to 30 contain all combinations possible (i.e. $3 \times 3 \times 1 \times 2 = 18$ combinations) of ARCs having a positive Δw_i value. However 9 Basic Sets found among Nos. 1 to 13 will be repeated once in the group 14 to 30. The actual number of Basic Sets considered for solution can be controlled by some pre-defined limit.

6. Generation of Daily Resource Allocation

The algorithm for resource levelling in multi-resource projects is fully described in [1], [19], and only a brief summary of the method is given here.

The starting point is the Basic Set in which each activity is performed by a specific ARC. It is assumed that scheduling is done on a daily basis, though any other convenient period would do. Each day, a list of candidate activities is prepared for the selection process. Selection is determined by an objective function called the schedule deviation (Δ), whose value is the weighted sum of squares of the deviations between the average future resource load and the currently assigned resource level

$$\text{i.e.} \qquad \Delta = \sum_{k=1}^{p} \mu_k \left[\frac{w_{kd}}{\lambda_d} - q_{kd} \right]^2 \qquad (8)$$

where:

w_{kd} is the remaining (i.e. unscheduled) work days for resource k, i.e. from the start of day d until the project completion.

q_{kd} is the amount of resource k allocated on day d.

λ_d is the remaining critical path duration, after activities on day d have been completed.

Thus the activity combination that minimizes the schedule deviation Δ is selected for day d's schedule. An enumeration procedure, based on binary numbers, is used for efficiently generating all activity combinations. With each enumeration, the values of the schedule deviation Δ is calculated and that combination yielding minimum Δ is selected for day d. Penalties for activity splitting and for critical activity delays are added to the appropriate objective function, i.e. schedule deviation, which is then minimized.

The levelling process progresses day by day until all activities have been completed, i.e. when $w_{kd} = 0$ for $k = 1, \ldots, p$ and $\lambda_d = 0$.

The main advantage of this levelling method is that the next periods work allocation depends only on the completed activities, i.e. it is not necessary to reschedule the entire remaining project in order to determine the next days work allocation.

7. Cost Calculations

Daily costs are found by using equations (1), (2), (3) and (4) which are added to give the accumulated cost of the project. The lower bound for total cost, using each basic set, is also calculated each day. These lower bounds are determined as follows:

The ratio w_{kd}/λ_d gives the average future workload for resource k. The future cost bound is calculated on the assumption that future resource allocations will be made by using ideal balanced profiles, and when added to the present accumulated cost, gives the lower bound for the total cost.

An $\bar{\lambda}$-element cost vector $\underline{C}_\lambda$ is defined, whose ℓth element ($\ell = 1, \ldots, \bar{\lambda}$) is the optimal cost for a project duration of $\lambda = \lambda_1 + \ell - 1$. Preliminary values of $\underline{C}_\lambda$ are ∞, for $\ell = 1, \ldots, \bar{\lambda}$.

To save on computation time, the lower bounds are compared with the appropriate elements of the cost vector $\underline{C}_\lambda$, which, if greater or equal (to $\underline{C}_\lambda$) results in the cancellation of the current basic set.

Thus, when $\lambda_d = 0$, the accumulated cost is minimal for the appropriate project duration and its value is entered into $\underline{C}_\lambda$.

8. The Algorithm

The algorithm is designed to select the basic sets along the lines described in Section 5. For each basic set, the following procedure takes place:

Resources are allocated on a daily basis, using the binary method for the enumeration of all assignable activity combinations. Penalties for splitting activities as well as for delaying critical activities are included, and finally that combination of activities giving minimal value of Δ is adopted for the daily schedule. Daily accumulated costs and cost bounds are calculated, and a comparison is made between the cost bound and the current minimum cost for the same duration. This procedure is repeated until all activities have been allocated. The flow chart for the algorithm is given in Fig. 5.

The algorithm has been tested on several 50-activity projects having 3 resource types, an average of 3 ARCs per activity and 5 to 150 Basic Sets. Computation times ranged from 200 sec. to 37 min. on a CDC 3600. It is expected

that this computer could handle projects in the order of 200 activities, using 5 resources, an average of 3 ARCs per activity and about 15 Basic Sets, for about a half hour computation time.

9. Sensitivity Analysis

A sensitivity analysis of the model was made with several major variables and the results are fully documented in [19]. The main results are summarized:

No.	Factor	Range Tested	Remarks
1	Activity Splitting	0 to ∞	Maximum increase in total cost under 2%
2	Penalty for Delays of critical activities	0 to ∞	Maximum increase in total cost under 2%
3	Number of Basic Sets	50 to 150	Maximum cost reduction of 0.2%. Computer time 16m to 37m on CDC 3600
4	Order Strength*	0.2 to 0.9	Computer time from 200 sec. to 380 sec. for 5 Basic Sets (50 activity project) using a CDC 3600.
5	Duration of critical activity	6 to 3	Project duration decreases. Total cost increases by 1 ½%.
6	Duration of non-critical activities	7 to 13	Project duration increases. Total cost increases by 1 ½%.
7	Resource Quantities	2 to 20 for *one* resource	Project duration increases. Total cost increases by 1½%

*Order Strength = $\dfrac{\text{No. of nonzero cells in minimum precedence matrix}}{\text{total number of cells in minimum precedence matrix}}$

10. Applicability of the Algorithm

10.1 The input data is used in a form that is immediately available to a project co-ordinator or manager. The use of ARCs encourages the planner to search out alternative ways of executing activities, a task which, in itself, is likely to uncover unusual but efficient modes of operation. Thus, the effect of compressing an activity is simply achieved through specifying as many ARCs as needed to explore the region of compression.

Cost data on resources are also readily available in its practical form—through the company cost records (when using one's own resources) and through contractor quotations when resources are to be supplied. Perhaps the Hiring and Dismissal Costs are generally not immediately available, but no serious difficulty is envisaged in estimating them.

10.2 Project control is simplified: it is done periodically (as with common practice), but the following period's schedule is obtained without being forced into rescheduling the entire project. This means lower computation costs, but more important, the next period's schedule is unaffected by any error that may occur (unintentionally) in the specification of some future activity.

10.3 Experience with this algorithm indicates that good solutions are obtained for the total project variable costs after the generation of only a few Basic Sets. The sensitivity analysis gave a maximum increase of 2% in these costs over wide ranges in the variables.

Generating the traditional "project variable cost-time" trade-off curve, immediately leads to the least cost solution for the project—something that no other practical resource-constrained project scheduling algorithm is able to accomplish.

Conclusions

The paper shows that this new approach in obtaining project time-cost trade-off relationship is feasible. The model uses realistic input data, which permits the inclusion of several resource combinations for each activity, the hiring and dismissal costs of resources, and the resources themselves are ranked in order of importance. The data also requires the inclusion of resource cost functions, which can be both linear and non-linear.

Penalties for delaying critical activities and for splitting activities are used in the heuristic algorithm to generate optimal project schedules for a wide range of project durations from which the conditions for least total cost can be found.

Sensitivity analysis shows that the algorithm is effective in tackling a wide range of project problems which differ in size (resource types and Basic Sets), order strength, the number and amount of combinations of resources and also in cost characteristics.

Most important is the fact that the algorithm is easy to use. Daily scheduling of activities depends only on completed activities and does not depend on the complete scheduling of the remaining project, thus, providing a realistic aid in project planning and control.

List of Abbreviations

C_λ – Optimal cost vector

d – Day number

g_{1k}, g_{2k} – Fixed and variable hiring cost for the k*th* resource type.

h_{1k}, h_{2k} – Fixed and variable dismissal cost for the k*th* resource type.

i – Activity number $(i = 1, \ldots, n)$

j – ARC index $(j = 1, \ldots, L_i)$

k – Resource type index $(k = 1, \ldots, p)$

ℓ – Optimal cost vector component index ($\ell = 1, \ldots, \bar{\lambda}$)

L_i – Number of alternatives of the i*th* activity

m – k*th* resource type curve sector index

n – Total number of project activities

p – Total number of resource types

g_{kd} – k*th* resource amount working on day d

r_{mk} – m*th* break point on the k*th* resource cost curve

u_{mk} – k*th* resource unit cost in the m*th* sector on resource cost curve

w_i – Weighted workload for activity i

w_{kd} – Remaining work from the beginning of the day d until the project completion

x_1, x_2 – Binary variables

Δ – Schedule deviation

$\bar{\Delta}$ – Optimal value of schedule deviation

λ – Project duration

λ_1 – Minimal critical path length

λ_2 – Maximal critical path length

λ_d – Updated remaining length of critical path on day d

$\bar{\lambda}$ – Range of project durations.

References

[1] Dar-El, E. M. and Tur, Y., "An Algorithm for Resource Levelling of a Multi-Resource Project," *Operations Research, Statistics and Economics,* Mimeograph Series No. 157, Technion, I. I. T., Faculty of Industrial and Management Engineering, Haifa, Israel.

[2] Davis, E. W. and Heidorn, G. E., "An Algorithm for Optimal Project Scheduling under Multiple Resource Constraints," *Management Science,* **17**, 12, B803-B816 (1971).

[3] Davis, E. W., "Project Scheduling under Resource Constraints–Historical Review and Categorization of Procedures," *AIIE Transactions,* **5**, 297-313 (1973).

[4] Davis, E. W., "Project Network Summary Measures Constrained–Resource Scheduling," *AIIE Transactions,* 7, 2, 132-142 (1975).

[5] Davis, E. W. and Patterson, J. H., "A Comparison of Heuristics and Optimum Solutions in Resource-Constrained Project Scheduling," *Management Science,* **21**, 8, 944-955 (1975).

[6] Gonguet, L., "Comparison of Three Heuristic Procedures for Allocating Resources and Producing Schedule," in *Project Planning by Network Analysis,* Lombaers, H. J. M., Ed., North Holland, Amsterdam (1969).

[7] Hastings, N. A. J., "On Resource Allocation in Project Networks," *Operational Research Quarterly,* **23**, 2, 217-221 (1975).

[8] Herroelen, W. S., "Resource Constrained Project Scheduling–The State of Art," *Operational Research Quarterly,* **23**, 2, 261-275 (1972).

[9] Hooper, P. C., "Resource Allocation and Levelling," *Proceeding of 3rd C.P.A. Symposium,* Operations Research Society, London (1965).

[10] Johnson, T. J. R., "A Branch and Bound Algorithm for the Resource Constrained Project Scheduling Problem," *Research Report, M.I.T.,* (1967).

[11] Lambourne, S., "Resource Allocation and Multiproject Scheduling (RAMPS)–A New Tool in Planning and Control," *The Computer Journal,* **5**, 300-304 (1963).

[12] Mason, A. T. and Moodie, C. L., "A Branch and Bound Algorithm for Minimising Cost in Project Scheduling," *Management Science,* **18**, 4, B158-B173 (1971).

[13] Pascoe, T. L., "Heuristic Method for Allocating Resources," Thesis for PhD, University of Cambridge (1965).

[14] Patterson, J. H., "Alternate Methods of Project Scheduling with Limited Resources," *Naval Research Logistics Quarterly,* **20**, 4, 767-784 (1973).

[15] Petrovic, R., "Optimization of Resource Allocation in Project Planning," *Operations Research,* **16**, 3, 559-567 (1968).

[16] Prabhakar, T., "Resource Allocation/Levelling in Engineering/Construction Projects," *Paper Presented at the 37th Meeting of ORSA,* 20-22 (April 1970).

[17] Pritsker, A. A. B., Watters, L. J. and Wolfe, P. M., "Multiproject Scheduling with Limited Resources; A Zero One Programming Approach," *Management Science,* **16**, 1, 93-109 (1969).

[18] Sunaga, T., "A Method of the Optimal Scheduling for a Project with Resource Restrictions," *Journal of Operations Research of Japan,* **12**, 2, 52-64 (1970).

[19] Tur, Y., "Optimizing Resource Costs when Specifying the Project Completion Date," PhD Dissertation, Technion–I.I.T., Haifa, Israel (1975).

[20] Wiest, J. D., "Some Properties of Scheduling for Large Projects with Limited Resources," *Operations Research,* **12**, 3, 395-418 (1964).

[21] Wiest, J. D., "A Heuristic Model for Scheduling Large Projects with Limited Resources," *Management Science,* **13**, 6, B359-B377 (1967).

[22] Zaloom, V., "On the Resource Constrained Project Scheduling Problems," *AIIE Transactions,* **3**, 4, 301-305 (1971).

Dr. Yigal Tur is Engineering Project Manager, Nuclear Research Center, Dimona, Israel. He manages Large-scale engineering projects and industrial applications of operations research. Dr. Tur received his BS, MS and PhD degrees from the Technion.

Dr. Ezey M. Dar-El holds the Harry Lebensfeld Chair in Industrial Engineering at Technion–Israel Institute of Technology, Haifa, Israel. His current research involves design of production systems, work measurement, design of work systems and human factors aspects of Industrial Safety. Dr. Dar-El is Director of the Research Center on Human Factors Engineering and Work Safety and is a Director of the Productivity Institute of Israel. He holds a BME and PhD from the University of Melbourne, Australia. Dr. Dar-El is a Senior Member of AIIE.

Synthesizing Short Run Production and Marketing Decisions

ULF PETER WELAM
Boston University School of Management *
685 Commonwealth Avenue
Boston, MA 02215

Abstract: Marketing mix decisions influence the demand for a firm's products, but most production-smoothing models treat forecast demand as an exogenous quantity. Here a model for evaluating the benefits of a closer coordination of short-run production and marketing decisions is presented. Ease of implementation in a typical functionally organized firm was a key consideration in the specific design of this model. Numerous experiments have been run and a few typical results are discussed in this paper.

■ Most manufacturers experience fluctuating demands for their products. To reduce the costs associated with such demand fluctuations, the marketing department attempts to stimulate sales during slack periods, and the production department pays overtime to its workers, hires and lays off workers, and allows its level of inventory to fluctuate in order to absorb the swings in demand. There is usually strong interactive effects between these kinds of managerial options in responding to demand fluctuations. Appropriately mixing these options is typically referred to as aggregate planning or, alternatively, production smoothing. In the management science literature, a rather formalized view of this problem has evolved, where aggregate planning or production smoothing is considered to be the determination of an optimal response to demand fluctuations by simultaneous selection of production rate, size of work force, inventory level and amount of subcontracting (if available) for each period in some finite planning horizon, the length of which is commonly of an order of magnitude of 6 to 12 periods (months).

In 1938 Erich Schneider [26] formally analyzed the production smoothing problem, but not until the 1950's did management scientists seriously address this issue. The work by Holt, Modigliani, Muth and Simon (HMMS) [11] has undoubtedly been the major stimulus in this regard. Following their lead, many authors have put forth a variety of model formulations and solution methods, [3, 4, 10, 13]. An article by Silver [27] summarizes the major issues and ideas involved.

Hitherto proposed solutions have, with few exceptions, treated demand as an exogenous variable, which can be predicted, but not influenced. In essence, analysis has been restricted to the production system, with the demand forecast as a major but exogenous system input. This sequential approach can produce an optimal solution only if the demand forecast *happens* to be for optimal demand levels.

Peterson extended [19] the analysis of the production smoothing problem by relaxing the assumption that actual shipments are identically equal to ordered shipments in all periods; the manufacturer can smooth distributor orders by over or undershipping. Within the planning horizon cumulative orders must equal cumulative shipments, but only first period decisions are implemented, so this balancing of total orders and total shipments need never be actually realized. A linear decision rule approach is retained, by assuming that the cost of over- or undershipping in any period is quadratic in the cumulative discrepancy between shipments and orders. Peterson's model is an interesting and significant contribution, but there appears to be

Received August 1975; revised February 1976; August 1976.

*Currently at Boston University–Brussels, 8 Avenue Roger Vandendriessche, B-1150, Brussels, BELGIUM.

a number of unresolved issues regarding its practical implementation.[1]

Tuite's [33] model deals explicitly with demand influencing activities. Demand is redistributed over the planning horizon through the use of a seasonal discount schedule. Assuming that the same discount rate is to be used for each month in the planning horizon leads to a particularly simple model with only one demand smoothing variable, but the assumption appears to be unnecessarily restrictive. There is the further restrictive assumption that total demand over the horizon is unaffected by the seasonal discounts and that a demand increase in one single period must be exactly balanced by a demand decrease in some other period.

The most comprehensive model published so far is one by Damon and Schramm [6] which deals with integration of production, marketing and finance decisions.

A direct causal relationship exists between the bulk of the production smoothing costs typically incurred by a firm and the swings in demand it is confronted with, so that by giving explicit consideration to demand smoothing activities a firm should be able to further reduce the minimum costs it must incur when relying solely on production smoothing measures

Moreover, in situations where managerial discretion with respect to traditional production smoothing variables is significantly reduced, for instance by union agreements, the conditions of the labor market in general etc., demand smoothing rather than production smoothing may become the primary vehicle for reducing the costs attributable to demand fluctuations.

This paper studies the consequences of relaxing the assumption of an exogenous demand forecast by formulating a model which simultaneously considers both production smoothing and demand smoothing activities and provides a logical framework for analyzing the potential benefits from coordination of certain production and marketing decisions. Total minimum cost is studied in two different situations: (i) simultaneous determination of marketing and production decisions (PD-model) (ii) sequential decision making, i.e., first the optimal marketing mix is determined, (D-model); this fixes the demand forecast which serves as an input to the problem of minimizing the total cost of smoothing production, (P-model).

The P-Model (Production Smoothing Model)

The general approach of this paper is independent of the particular type and functional form of the cost relationships included in the production smoothing model. To help focus on the central issue of this paper–coordination of decision making in the marketing and production subsystems–this latter system is represented by the HMMS model, which should be familiar to most readers.

The cost components of the P-model are regular payroll costs, hiring and layoff costs, overtime costs and inventory costs, and the model is specified as

$$\min \sum_{t=1}^{T} C_t \tag{1}$$

and

$$C_t = [c_1 w_t \quad \text{regular payroll costs} \tag{1-1}$$

$$+ \; c_2(w_t - w_{t-1})^2 \quad \text{hiring and lay off costs} \tag{1-2}$$

$$+ \; c_3(p_t - c_4 w_t)^2 + c_5 p_t - c_6 w_t \quad \text{overtime costs} \tag{1-3}$$

$$+ \; c_7(z_t - z_t^*)^2] \quad \text{inventory-connected costs} \tag{1-4}$$

$$z_t^* = c_8 + c_9 d_t \quad \text{optimal inventory level} \tag{1-5}$$

subject to the restrictions for $t=1, \ldots T$ that (1-5)

$$z_t = z_{t-1} + p_t - d_t \tag{1-6}$$

where

C_t	total incremental cost in period t
p_t	production rate in period t
w_t	work force level in period t
z_t	inventory level in period t
z_t^*	optimal inventory level in period t
d_t	forecasted demand for period t
c_i $i=1,\ldots 9$	empirically estimated parameters
T	number of periods in the planning horizon

The D-Model (Demand Smoothing Model)

The demand smoothing system is conceptualized as encompassing those aspects of the total marketing system which are of particular significance for the relatively short-run production and marketing decisions of interest here. It is reasonable to assume that, in the short run, price and sales promotion are the primary means through which the firm can effect changes in demand. Sales promotion is here taken to be an aggregate measure of the total promotional activity in the firm and is stated in terms of its dollar value. In specifying the D-model special attention was given to two considerations regarding the implementation of the PD-model in an actual environment.

Firstly, price and sales promotion are both important key variables in the larger total marketing system. Decisions

[1] The game theoretic aspects of the process of negotiating the size of the side payments to be made to the distributors as a compensation for their absorbing part of the manufacturer's smoothing costs have been commented upon by Eilon [5].

with respect to these variables will never be made with reference only to such subsystems as the production smoothing and demand smoothing systems, but will normally involve consideration of a great many additional factors beyond those being given explicit treatment here. It is highly desirable that the structure of the model reflects this point of view, i.e., that in a real world application the PD-model will necessarily only be part of a larger whole.

Secondly, the PD-model cuts across current traditional functional areas of management. In principle, most managers will probably readily agree with Churchmann [5] that "the systems approach is not a bad idea," but genuine mangerial acceptance of the full implications of such a systems approach might be much more difficult to secure. It is, in any event, clear that the implementation of a model which does not immediately fit into any one particular functional area will require a special effort on the part of management.

These kinds of considerations suggest that in attempting to formulate an *implementable* model, it must be recognized that marketing decisions are frequently the exclusive prerogative of marketing management, and the same goes for production decisions and production management. On the other hand, and equally important, a major potential benefit from implementing the model is that it can become *instrumental* in furthering the coordination between the activities of the departments. The goal therefore is to construct a model consonant with the decision making process in a typical functionally organized company, but which also provides a linkage between the production and marketing departments, whereby tradeoffs between the two may be evaluated.

To put these ideas into sharper focus, assume that the total marketing system is described by the following variables:

pe_t	price in period t
sp_t	sales promotion in period t
m_t	all other marketing variables in period t
ce_t	production cost in period t
r_t	revenue in period t
d_t	forecasted demand for period t

where

$$d_t = f(pe_t, sp_t, m_t, pe_{t-1}, sp_{t-1}, m_{t-1}, \ldots, pe_o, sp_o, m_o) \quad (2\text{-}1)$$

In formulating an objective function for the marketing system, it seems reasonable to assume that while marketing management does not ignore production costs, they are accounted for in rather summary fashion by using a constant marginal cost, ce_t, to represent period t production costs. Letting $c(m_t)$ symbolize the costs of all marketing instruments other than price and promotion, the net revenue function in period t is

$$r_t = (pe_t - ce_t)\, d_t - sp_t - c(m_t) . \quad (2\text{-}2)$$

The objective function for the total marketing system is

$$\max \sum_{t=1}^{T} r_t . \quad (2)$$

Letting superstars denote optimal values, one has

$$\sum_{t=1}^{T} r_t^* = \sum_{t=1}^{T} [(pe^*_t - ce_t)d_t^* - sp_t^* - c(m_t^*)] . \quad (2^*)$$

By assuming that this optimal solution is known, the PD-model can be put to as severe test as possible regarding its ability to find a solution which will improve upon that found from separately optimizing the marketing and the production smoothing systems and disregarding possible tradeoffs between the two. If to begin with, only a feasible solution to the marketing system is known, potential benefits from such tradeoffs will quite naturally increase.

For notational convenience, the variables of the D-model are expressed in terms of deviations from their optimal values vis-a-vis the total marketing system. (In case only a feasible solution is known, the deviations are of course measured from these feasible values). That is, for t=1, ...,T

$$x_t = pe_t - pe_t^*$$

$$y_t = sp_t - sp_t^*$$

$$\Delta d_t = d_t - d_t^*$$

$$\Delta r_t = r_t - r_t^* .$$

Assuming that all marketing instruments other than price and promotion do not change, and that over the range of values considered demand (2-1) is separable in (pe_t, sp_t) and m_t, then incremental revenue in period t is

$$\Delta r_t = (d_t^* + \Delta d_t)\,[(pe^*_t - ce_t) + x_t] - (sp_t^* + y_t) - [d_t^*(pe^*_t - ce_t) - sp_t^*] . \quad (3)$$

The change in demand depends on current price and promotion changes and possibly on past such changes

$$\Delta d_t = h(x_t, y_t, x_{t-1}, y_{t-1}, \ldots, x_o, y_o) . \quad (4)$$

To simplify notation let $\pi_t^* = pe_t^* - ce_t$, so that (3) becomes

$$\Delta r_t = d_t^* x_t + (\pi_t^* + x_t)\,\Delta d_t - y_t .$$

To pursue the analysis further it is necessary to specify the relationship $\Delta d_t = h(x_t, y_t, x_{t-1}, y_{t-1}, \ldots, x_o, y_o)$. In a real world application, the question as to what specific functional form for h is most suitable, must of course be resolved in light of a careful analysis of the particular characteristics of the problem at hand. While it would carry us too far afield to elaborate on this issue here, it should be pointed out that the method for solving the PD-model is relatively independent of the specific functional form chosen. The form for h used in the remainder of this paper

is representative of currently available theories and empirical evidence from marketing research [1, 2, 7, 15-18, 21, 23-25, 28-31, 34, 35]. These suggest that current demand for a product is influenced not only by current marketing activities but also by past ones. One can conceptualize a process where the marketing activities of a firm over time build up a form of capital asset often referred to as good will [16], whose current value in conjunction with current price and promotion determines the current level of sales. For purposes of empirical estimation the concept of goodwill must be made operational and a common proxy measure for this goodwill is past demand. For example, in postulating a sales (S_t) to adversiting (A_t) relation of the form

$$S_t = a_0 + a_1 S_{t-1} + a_2 f(A_t).$$

Simon [31, p. 30] writes: "If one makes some frequently reasonable assumptions, one can think of last period's sales as being the embodiment of the residual effect of all advertising up to the last period. And the same assumptions suggest that a certain proportion of last period's sales would carry over into this period. This is also consistent with the idea that last period's sales represent a stock of purchase habits, a large part of which would still be in effect in this period even if no advertising were done this period. Following on the above, one may use last period's sales as a variable in the regression, as a proxy for the effect of advertising in all prior periods." Alternatively one can start with a relation of the form

$$S_t = a_0 + b_0 f(A_t) + b_1 f(A_{t-1}) + \cdots b_i f(A_{t-i}) + \cdots .$$

Under suitable assumptions this relation can for estimation purposes be transformed into the so-called distributed lag form represented by the equation immediately above. Chapter 5 of Kotler [14a] gives a good summary and discussion of some alternative formulations and transformations that can be used.

Assume then, that for each period in the planning horizon changes in price and sales promotion affect demand according to the distributed lag model:

$$\Delta d_t = \Delta d_t^1 + \Delta d_t^2 \tag{6}$$

$$\Delta d_t^1 = \alpha_t x_t + \lambda \Delta d_{t-1}^1 \tag{6-1}$$

$$\Delta d_t^2 = \beta_t y_t - \gamma_t y_t^2 + \rho \Delta d_{t-1}^2 \tag{6-2}$$

where Δd_t is the total change in demand in period t and is assumed to be the sum of the change attributable to price changes (Δd_t^1) and promotional changes (Δd_t^2), α_t, β_t, γ_t, λ, ρ are estimated parameters, and $\alpha_t < 0, \beta_t, \gamma_t > 0$. The parameters[2] λ and ρ represent intertemporal carry over effects.

A positive ρ represents the traditional marketing assumption that sales in any one period is influenced not only by the promotional efforts during that same period but also by the promotional efforts in all previous periods. This accumulated promotional capital deteriorates at a rate of r per period, and hence the net interperiod carry over effect is $1 - r = \rho$.

A number of authors claim that in many instances a sales increase in the current period attributable to a stepped up promotional activity is nothing but a borrowing of sales from the future, thereby implying a negative ρ. Both cases can be accommodated within the framework of the model. Similarly, the carry over effect of price variations can be either positive or negative. The demand forecasts $\{d_t^*\}$ is influenced by a great many factors, e.g., the general state of the economy, the competitive situation, the marketing mix, etc. If this demand forecast represents an equilibrium position, then it seems reasonable that the incremental demand that can be generated in the current period by a price reduction will in part stem from an earlier placement of orders by the distributors and that such an earlier placement will occur if the marginal revenue accruing to the distributor from the price reduction is at least as large as his marginal inventory carrying costs. Conversely, an increase in price will cause a postponement of distributors' orders. On the other hand, there may well be situations where price variations produce exactly opposite results, e.g., lowering the price of the product now may actually stimulate increased usage of it in the future. Expanding (6-1) and (6-2) gives

$$\Delta d_t^1 = \sum_{i=1}^{t} \lambda^{t-i} \alpha_i x_i + \lambda^t \Delta d_0^1 \tag{7-1}$$

$$\Delta d_t^2 = \sum_{i=1}^{t} \rho^{t-i} (\beta_i y_i - \gamma_i y_i^2) + \rho^t \Delta d_0^2 \tag{7-2}$$

The values of the initial conditions Δd_0^1 and Δd_0^2 will influence the specific magnitude of the optimal values for the decision variables. The subsequent development is, however, completely independent of the values of Δd_0^1 and Δd_0^2 in a specific instance. For notational convenience, assume therefore that $\Delta d^1 = \Delta d^2 = 0$. The minor changes which occur when this is not the case are completely straightforward. Then substitution of (7-1) and (7-2) into (5) gives

$$\Delta r_t = d_t^* x_t + (\pi_t^* + x_t) \left[\sum_{i=1}^{t} \lambda^{t-i} \alpha_i x_i + \sum_{i=1}^{t} \rho^{t-i} (\beta_i y_i - \gamma_i y_i^2) \right] - y_t \tag{8}$$

The objective function for the D-model can then be written as

[2] $\gamma_t > 0$ is reasonable of course only with respect to a specified limited range of y_t.

$$\max \sum_{t=1}^{T} \Delta r_t = \max \sum_{t=1}^{T} \left\{ d_t^* x_t + (\pi_t^* x_t) \right.$$

$$\left. \cdot \left[\sum_{i=1}^{T} \lambda^{t-i} \alpha_i x_i + \sum_{i=1}^{t} \rho^{t-i} (\beta_i y_i - \gamma_i y_i^2) \right] - y_t \right\}. \quad (9)$$

To facilitate a clear cut comparison between the simultaneous and sequential models, it will be assumed from here that x and y measure the deviations from those price and promotion levels which are optimal for the marketing system when production smoothing costs are ignored. Consequently, if the D-model itself is maximized, the optimal solution should be $x_t^*, y_t^* = 0$ for all t, and the value of the objective function (9) should then be zero. That is, if price and promotion adjustments are evaluated only in terms of their impact on net revenue changes exclusive of production smoothing costs, then by assumption it is optimal not to adjust at all.

The PD-Model

For the PD-model, the problem is to find the values of the decision variables that minimize the sum of the production smoothing costs and the reduction in "marketing net revenue," i.e., find values of p_t, w_t, z_t, x_t, y_t which satisfy

$$\min \sum_{t=1}^{T} (C_t - \Delta r_t) \quad (10)$$

$$\text{s.t. } z_t = z_{t-1} + p_t - (d_t^* + \Delta d_t). \quad (11)$$

Experimental Assumptions and Design

The performance of a hypothetical firm was studied under two alternative assumptions.

1. That the demand forecast is exogenous, i.e., it cannot be influenced by the decision maker. This assumption is represented by the P-model, so the problem is to find optimal values for production, workforce and inventory, for each period in the planning horizon. First period decisions are recorded as implemented decisions.
2. That the demand forecast is partly endogenous, i.e., it can be influenced by the decision maker by allowing price and sales promotion to deviate from their optimal values vis-a-vis the marketing system. This assumption is represented by the PD-model. Here the problem is to find optimal values for production, workforce, inventory, price and sales promotion for each period in the planning horizon. First period decisions are recorded as implemented decisions.

The performance of the firm under each of the two listed assumptions was studied over a time span of I periods. In each period a planning horizon of T periods is used and the decisions for the first of these T periods are recorded as implemented decisions. The model is then advanced one period by using the implemented decisions for period i as initial conditions for period $i + 1$. In going from period i to $i + 1$, the demand forecast is updated by shifting it back one period and by reading in a new value for the last period (T) in the planning horizon. That is, a rolling horizon was employed in studying the performance of the indicated systems.

The demand and cost data for the paint factory example described by HMMS [9] have been used by several other authors, e.g., [14] and [16] as a basic reference point for evaluating the performance of their proposed models. This has the obvious advantage of facilitating comparisons between alternative models, and in keeping with this "tradition," this study used the paint factory data to the fullest possible extent. However, for the marketing parameters of the D-model it was necessary to assign values that were deemed reasonable in light of current marketing research. In order to retain as much commonality as possible with respect to the input data used in the experiments reported here and in those reported by others, it was assumed that the demand data for the paint factory represents a forecast which is optimal for the marketing system when considered by itself. That is, for the sequential model it is assumed to be the forecast resulting from optimizing the marketing mix before turning to the question of minimizing the production smoothing costs. This experimental assumption drastically reduces the "degrees of freedom" available in assigning values to the parameters of the marketing system and should therefore also drastically reduce the possibility of any accidental systematic bias in the experimental results. The details on this reduction in the degrees of freedom are discussed in a working paper [36] available from this author.

Model Solution

Each of the above steps in the simulation involves finding the optimal solution to a nonlinear programming problem, and some strategies for accomplishing this will now be sketched. For a detailed discussion of this topic the reader is referred to the above working paper [36].

The P-model can be transformed into an unconstrained quadratic function in 2T variables by using the inventory balancing Eq. (1-6) to eliminate the inventory variables. It is certainly reasonable to assume that c_2, c_3, c_4, and c_7 are all nonnegative, in which case the quadratic part of the cost function is positive definite. Thus computing the *unique* optimal solution to the P-model essentially involves inverting a matrix of order 2T. Furthermore this matrix has a very special structure which allows for a significant reduction in computational requirements, if one is willing to write a special purpose code as opposed to simply using a general purpose matrix inversion program. See [37] for further details on this.

The PD-model is transformed into an unconstrained nonlinear function in 4T variables (production, workforce, price changes and promotion changes) by again using the inventory balancing equation to eliminate the inventory variables. If this function is convex, then any local minimum is a global minimum. Unfortunately the PD-function is so complex that deriving closed form expressions for the principal minors of its second derivative matrix is very complicated and tedious, and so far this has been done only for T=1 and T=2. In these instances it is clear though that the PD-function is convex for reasonable parameter values, i.e. values such that production smoothing costs are positive, demand is a decreasing function of price, and demand exhibits decreasing returns to scale (saturation effects) with respect to promotion. For ($T > 2$ numerical experiments were relied on to assert the convexity or nonconvexity of thePD-function. A great many different starting points were tried for the iterative search method used to optimize the PD-function, and in all instances the same solution was obtained. This does not conclusively prove that the PD-function is convex for "reasonable" parameter values, but the numerical experiments in conjunction with the analytic results for $T > 2$ give some reassurance that the PD-function does possess a unique optimal solution for parameter values of practical interest.

Specifically, the numerical results discussed in the next section were obtained as follows: The P-model was solved by simply postmultiplying the inverse of its second derivative matrix by its first derivative vector. This yields the optimal production and workforce levels and the optimal inventory levels are then obtained directly from the inventory balancing Eqs. (1-6). The PD-model was solved using a conjugate direction iterative search method first presented by Powell [20] and later elaborated on by Zangwill[38]. This method is attractive in an experimental context in that it does not require calculation of any derivatives. The robustness of the PD-model could therefore be tested very quickly since restructuring it involved very little programming effort. However if the model is to be run on a routine periodic basis a second order method such as the Newton-Raphson algorithm [9] is probably preferrable in that it will substantially reduce computer run time needed.

From the user's point of view the essential trade off is between the coding setup cost and the cost of running the program. A computer program embodying a second order optimization method is necessarily more problem dependent than a program embodying a first or zero order method (i.e. no derivatives required), and therefore requires substantial programming on the part of the user.

Twelve months of factory operations were simulated. In each period a six-month planning horizon was used, thus requiring the optimization of a 12-dimensional quadratic function and a 24-dimensional nonlinear (probably convex) function. The optimization algorithms were coded in Fortran IV and on an IBM 360/50 the total CPU-time required for a twelve period simulation ranged from eight to ten minutes.

These run times are not excessive, but in iterative nonlinear optimization, solution times tend to increase at a far greater than linear rate in the number of variables. Thus a 12 month as opposed to a 6 month planning horizon would probably result in solution times of an order of magnitude of 5-6 minutes for each simulated period. Decomposition can be used to overcome this dimensionality problem, and in the appendix it is shown how the iterative numerical optimization can be reduced to a T-dimensional problem. Computation time will be cut drastically, but the amount of computer coding will naturally increase. This is the basic trade off decision that a model user must make.

Some Experimental Results

With respect to the input data shown in Exhibit 1, it should be noted that Δd_0^1 and Δd_0^2 were set to zero only for the first period in the simulation. For the remaining periods Δd_0^1 and Δd_0^2 were assigned the values of Δd_1^1 and Δd_1^2 for the immediately preceding period. I.e., the runs simulated a system with a "rolling horizon." Arbitrary, but of course reasonable, values were chosen for the intertemporal carry over parameters, λ and ρ, and for the optimal gross margin π^*. The vectors of demand parameters α and β are now uniquely determined by the assumption that the demand forecast results from separately optimizing the marketing system. Monthly sales pattern is the exogenous demand forecast for the P-model. With reference to the PD-model it is to be interpreted as the demand forecast which results from optimizing the total marketing system. In the runs discussed here only the first 17 entries were used, since 12 months of firm operation was simulated using a rolling planning horizon (T) of 6 months.

Exhibit 1: Input data for simulation runs.[a]

Monthly sales Pattern:

Per	1- 6:	430,	447,	440,	316,	397,	375
"	7-12:	292,	458,	400,	350,	284,	400
"	13-18:	483,	509,	500,	475,	500,	600

Initial inventory (z_0): 263

Initial workforce (w_0): 81

Cost Parameters (c_i)

c_1 - c_5: 340.00, 64.30, 0.20, 5.67, 51.20

c_6 - c_9: 281.00, 0.0825, 320.00, 0.0

Optimal gross margin (π^*): 200.00

Initial demand deviations($\Delta d_0^{(1)}$, $\Delta d_0^{(2)}$):0, 0

Intertemporal carry over parameters

λ = 0.70

ρ = 0.40

Demand parameters

$\alpha_1, \ldots, \alpha_6$: -1.86, -1.93, -1.92, -1.33, -1.89, -1.25

$\beta_1, \ldots, \beta_6$: .00201, .00202, .00205, .00214, .00238, .00333

$\gamma_1, \ldots, \gamma_6$: .000001

[a]The reader will note that wherever possible the input data used is taken from HMMS [6].

Exhibit 2: Totals of production (p), workforce (w) and inventory (z) decisions.

		p(units)	w (men)	z(units)
1	P	4,710	775	3,919
	PD	4,249	713	3,916
2	P	4,763	837	3,892
	PD	4,450	812	3,810
3	P	4,619	846	4,222
	PD	4,628	841	3,882

Exhibits 2, 3 and 4 show part of the output from the P and PD-model respectively, using the input values from Exhibit 1. Exhibits 2 and 3 show totals of implemented decisions and their cost consequences during the 12 months of simulated factory operation. The leftmost numbers refer to the three different simulation runs being discussed. P and PD identify the production and the production-demand smoothing models. The first four column labels in Exhibit 3 indicate cost of regular payroll, hire and fire, overtime and inventory. The total marketing revenue change from corrections to the "optimal" marketing decisons is denoted by ΔR. Finally, C_{tot} is total cost incurred during the 12 simulated months of factory operation.

Exhibit 3: Cost totals ($)

		C_{rp}	C_{hf}	C_{ot}	C_{iv}	ΔR	C_{tot}
1	P	263,330	4,828	25,717	1,325	-	295,200
	PD	242,394	7,275	18,097	952	-16,796	285,514
2	P	284,653	4,587	9,615	6,314	-	305,170
	PD	276,237	8,338	18	2,232	7,419	294,245
3	P	160,767	4,233	2,668	7,807	-	175,475
	PD	159,723	4,100	1,140	2,625	276	167,312

Exhibit 4 shows the actual deviations (x and y) in the simultaneous model from the "optimal" price and sales promotion levels in the sequential model and the corresponding change in the demand distribution (Δd) which were implemented during the 12 months of simulated factory operation.

Exhibit 4: Corrections to optimal marketing decisions r.

		1			2			3		
per	d*	Δd	x	y	Δd	x	y	Δd	x	y
1	430	-44	23	-102	-43	13	-172	5	-2	-81
2	447	-40	27	-132	-31	15	-88	-12	3	27
3	440	-42	28	-112	-26	12	-74	-12	5	22
4	316	-25	25	-95	-2	6	-15	10	-2	104
5	397	-38	26	-102	-11	3	-46	1	-2	30
6	375	-34	28	-91	-8	4	-50	0	0	31
7	292	-23	25	-90	-1	2	-51	17	-3	32
8	458	-56	32	-116	-30	8	-97	-12	2	-41
9	400	-43	34	-117	-24	12	-100	-4	3	38
10	350	-35	32	-128	-23	13	-129	18	-6	137
11	284	-24	26	-136	-17	12	-162	31	-19	31
12	400	-48	32	-144	-48	18	-286	25	-13	-167
		-455			-264			67		

Thus in the first example run, the P-model incurs a total cost for the 12 periods of 295,200 and the PD-model incurs 285,514. While the difference is not spectacular, it is still significant and it should be kept in mind that this latter solution represents an improvement over and above what is already in some sense an optimal solution. The result for the PD-model is arrived at by allowing a decrease in revenue of 16,796, which brings the production costs down to 268,718. Notice that a smoother demand pattern is accomplished by a decrease in demand in each period, with a proportionately larger decrease in the peak months. This demand decrease, which totals 455 units is effected primarily by a substantial price increase in each month. Thus for the given set of parameter values, a lower volume-higher price strategy is a more profitable alternative than the one found from sequential optimization.

A key factor here is the relatively high regular payroll-cost which in conjunction with the hiring and firing cost makes it unprofitable to completely satisfy the demand forecast, resulting from separately optimizing the marketing system. As is to be expected the demand decrease is largest for the peak months. This strategy leads to substantial revenue losses but this is more than compensated for by decreasing regular payroll and overtime costs.

In the second run the cost parameter for hiring and firing was increased to 664.3 (from 64.3). All other parameter values are the same as previously. Since it is now much more expensive to fire workers, a larger workforce is retained over the simulated 12 periods, and less overtime is used. Total cost for the P-model goes up to 305,170. (Exhibit 3.)

The solution to the PD-model indicates that a lower volume-higher price strategy is still preferable, but the price increases are now more modest and the cut in sales promotion has been reduced. Also changes in price and sales promotion are made more selectively, i.e., with greater differences between the changes in each period. The decrease in demand is now only 264 units and the revenue loss is 7,419. It is interesting to note that while previously it was profitable to lay off workers and assign overtime work to the remaining ones, this is no longer the case. Demand smoothing and laying off is now performed in such a way that virtually no overtime is used. Total incremental cost for the PD-model is 294,245, so the cost reduction is about the same as before, but the increased firing cost has caused a pronounced difference in the implemented decisions.

The third run used the parameter values in Exhibit 1 with the following modifications: c_1 = 190 (regular payroll) instead of 340 and z_0 = 363 (starting inventory) instead of 263. Total cost for the P-model is now 175,475, and for the PD-model 167,312.

A low volume-high price strategy is no longer the most profitable alternative. Instead, a sequence of increases and decreases in price and sales promotion brings about a smoothed pattern of demand together with a slight expansion of total demand for the simulated 12 periods. With a smoother pattern of demand, it is possible to maintain a

somewhat smaller workforce working less overtime, even though there has been a demand increase. The major part of the reduction in cost stems, however, from a reduction in inventory-connected costs.

The three runs just discussed are typical experimental results. In actual applications of the model, cost reductions will most likely be larger than those indicated by these experimental results. An optimal solution to the marketing system is rarely known, so coordinating marketing and production decisions is likely to be much more profitable than indicated by the preliminary experimental results presented here.

Concluding Remarks

A model extending the scope of traditional production smoothing anlaysis was formulated and experimented with. The model is not just a device providing optimal solutions to given problems, but more importantly, it assists management in formulating and answering "what if" type of questions. The computer implementation of this model therefore paid particular attention to flexibility and ease of use.

Extensions to several products and several different markets are interesting avenues for future research, but a main obstacle is the rapid increase in the number of decision variables. With a planning horizon of T months, P products and M markets, there will be $4T \times P \times M$ variables in the model. It is doubtful that currently available methods for nonlinear optimization can solve such large problems but semi-heuristic approaches [14] may still produce approximately optimal solutions even to very large problems. For example, one could construct a model similar to the one presented in this paper but with periods of unequal length in the planning horizon so as to reduce the size of the optimization problem.

In some lines of business, firmly established codes of "good business practice" may render a fully dynamic pricing policy "unethical," whereas a managerial commitment to a certain announced discount schedule over a prespecified number of time periods may be acceptable. A model which captures the essence of this situation is conceptually a straightforward generalization of the one discussed here. However, it then becomes necessary to consider not only the future effects of a price change in the current period, but also the effect on current period demand of all future price changes. Methodologically, the generalization is therefore not likely to be quite as straightforward, so here is another important area for future research on production and demand smoothing.

Appendix

It will be shown that the *numerical* optimization of the PD-model can be reduced to maximizing a function which depends on only T variables as opposed to the full set of 6T decision variables.

Matrix notation will be used in sketching the decomposition approach, for otherwise the notation becomes unbearably cumbersome. Subscripted lower case letters denote scalars, unsubscripted lower case letters denote vectors and upper case letters denote matrices.

Let w_o and z_o denote beginning workforce and inventory levels and let $q_0 = (T)c_7c_8^2 - 2c_2w_0$.

We adopt the convention that the elements of a vector are ordered with respect to time and to the extent possible use the same symbols as before. For example the production vector is denoted by p and defined as $p = (p_1, \ldots, p_T)$, and the price change vector is denoted by x and defined as $x = (x_1, \ldots, x_T)$. We also need the additional vector $y^2 = (y_1^2, \ldots, y_T^2)$, the summation vector $e = (1, \ldots, 1)$ and the ith unit vector $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$. For notational convenience the vector of demand deviations will be denoted by d_1, that is $d_1 = (\Delta d_1, \ldots, \Delta d_T)$.

Let $q_1 = (c_1 - c_6)e$ and $q_2 = c_5e$, $q_3 = -2c_7c_8e$.

I denotes an identity matrix of order T, 0 is a matrix whose elements are all zero and $Q_{11} = c_3I$, $Q_{12} = Q_{21} = -c_3c_4I$, $Q_{22} = c_2M'M + c_3c_4^2I$, $Q_{33} = c_7I$. Finally let,

$$M = \begin{pmatrix} 1 & & & 0 \\ -1 & \ddots & & \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{pmatrix} (T \times T) \text{ and } L = \begin{pmatrix} 1 & & & 0 \\ \vdots & \ddots & & \\ \vdots & & \ddots & \\ 1 & \cdots & \cdots & 1 \end{pmatrix} (T \times T)$$

so that $L = M^{-1}$

If demands were fixed the production department would proceed to minimize

$$F_1 = (p', w', z') \begin{pmatrix} Q_{11} & Q_{12} & 0 \\ Q_{21} & Q_{22} & 0 \\ 0 & 0 & Q_{33} \end{pmatrix} \begin{pmatrix} p \\ w \\ z \end{pmatrix} + (q_1', q_2', q_3') \begin{pmatrix} p \\ w \\ z \end{pmatrix} + q_o \tag{A-1}$$

$$\text{subject to } Mz = z_o e_1 + p - (d + d_1) \tag{A-2}$$

This is the P-model; its solution which is central to the subsequent development is

$$\begin{pmatrix} p^* \\ w^* \\ z^* \end{pmatrix} = \begin{pmatrix} p^o \\ w^o \\ z^o \end{pmatrix} + \begin{pmatrix} S_{11} \\ S_{21} \\ S_{33}\, M' \end{pmatrix} u^* \tag{A-3}$$

where $\begin{pmatrix} p^o \\ w^o \\ z^o \end{pmatrix} = -Q_q^{-1}$ i.e. the unconstrained solution to A-1, $S = Q^{-1}$ and S is partioned exactly as Q.

$u^* = D\ (d + d_1 - z_o e - p^o + M z^o)$ is the shadow price vector associated with A-2 and D is a positive definite matrix with very special structure.

Define the lower triangular matrices $\bar{A}$, $\bar{B}$, $\bar{C}$ whose elements are $\bar{a}_{i,j} = \alpha \lambda^{i-j}$, $\bar{b}_{i,j} = \beta \lambda^{i-j}$, $\bar{c}_{i,j} = \gamma \lambda^{i-j}$ for $i > j$ and zero otherwise.

If the marketing department acted independently it would

$$\text{Maximize } F_2 = d^* x + (\pi^* + x)' d_1 - e' y \qquad \text{(A-4)}$$

$$\text{s.t.} \qquad d_1 = \bar{A}x + \bar{B}y - \bar{C}y^2 \,. \qquad \text{(A-5)}$$

Simultaneous solution of the marketing and production sectors, that is solving the PD model requires finding those values of x, y, d_1, p, w, z which will

$$\text{maximize } F_2 - F_1$$

$$\text{s.t. (A-2) and (A-5).}$$

Note that the only decision vector common to both the marketing and production departments is d_1 which couples the two systems together by virtue of its precence in the inventory balancing Eq. (A-2). For any fixed d_1 the problem separates into two subproblems: Find values for x and y maximizing F_2 subject to (A-5), and find values of p, w, z minimizing F_1 subject to (A-2). The problem is of course that we don't know the specific d_1 for which optimizing the subproblems separately will yield the same result as optimizing them simultaneously. Our first step towards this end is to develop a closed form expression for the minimum production smoothing cost as a function of d_1.

In mathematical terms, we take the first partial derivatives with respect to p, w, z of $(F_2 - F_1)$ and use the necessary conditions for optimality to obtain an expression for the minimum production smoothing cost (call this F_1^*) as a function of d_1. The result is $F_1^*(d_1) = d_1' D\, d_1 - (z_o e_1 + d')\, D d_1$ where $D = [(1/c_3) I + (c_4^2/c_2) LL' + (1/c_7) MM']^{-1}$ which is a positive definite symmetric matrix of order T.

Next A-5 is used to eliminate x from explicit consideration in the numerical optimization. Thus, from A-5,

$$x = \bar{A}^{-1} d_1 - \bar{A}^{-1} \bar{B}y + \bar{A}^{-1} \bar{C}y^2 .$$

Conceptually it is possible to transform the model into one which is an explicit function of d_1 only, but in our particular case the mathematics simplify considerably if it is transformed in such a way that profit depends explicitly on y only.

For notational convenience we write (A-5) as

$$x = A\, d_1 - By + Cy^2 \,. \qquad \text{(A-6)}$$

The only nonzero elements of A, B, C are those on the diagonal and subdiagonal. Specifically $a_{ii} = (1/\alpha)$, $a_{i,i-1} = (\lambda/\alpha)$, $b_{ii} = (1/\beta)$,

$$b_{i,i-1} = (\rho/\beta),\ c_{ii} = (1/\gamma),\ c_{i,i-1} = (\rho/\gamma).$$

On substituting for x in F_2 we have

$$F_2 = (d^* + d_1)' (Ad_1 - By + Cy^2) + d_1' \pi^* - e' y \,.$$

Note that $F_2 - F_1^*$ is quadratic in d_1 so that the necessary optimality conditions with respect to d_1 yield a system of equations which is linear in d_1. Rearranging some terms we have

$$F_2 - F_1^* = d_1' (A-D) d_1 + d_1' (a + Cy^2 - By) + (d^*)' (Cy^2 - By)$$

where $a = A' d^* + D(z_o e_1 + d) + \pi^*$.

Setting the first partials with respect to d_1 equal to zero yields

$$2(A-D) d_1 + a + Cy^2 - By = 0$$

$$\text{or } d_1 = E(By - Cy^2 - a) \qquad \text{(A-7)}$$

where $E = (1/2)(A-D)^{-1}$.

Substituting for d_1 in $F_2 - F_1^*$, rearranging terms and calling the result F^* one now has

$$F^* = (y^2)' C' E C y^2 - 2 y' B' C y^2 + [2a' E + (d^*)'] C y^2 +$$

$$y' B' E B y - [2a' E + (d^*)'] B y - e' y + a' E a.$$

This is a fairly complicated function but there are now only T variables left to be determined numerically, and any of the iterative search methods discussed in the paper will easily handle this. Once the optimal promotion increments y^* have been determined by maximizing F^*, remaining quantities of interest can be computed from the formulas developed earlier. That is, d^*_1 is obtained from (A-7), next x^* is calculated from (A-6). Finally p^*, w^*, z^* are obtained from (A-3). Note here that u^* the shadow price vector associated with the inventory balancing equations depends on the vector of demand deviations d_1.

References

[1] Bass, F. M., "A Simultaneous Equation Regression Study of Advertising and Sales of Cigarettes," *Journal of Marketing Research* (August 1969).

[2] Bass, F. M., and Clarke, D. G., "Testing Distributed Lag Models of Advertising Effect," *Journal of Marketing Research* (August 1972).

[3] Bergstrom, G. L., and Smith, B. E., "Multi-Item Production Planning–An Extension of the HMMS Rules," *Management Science* (June 1970).

[4] Bowman, E. H., "Consistency and Optimality in Managerial Decision Making," *Management Science* (1961).

[5] Churchmann, C. W., *The Systems Approach*, Dell Publishing Company (1968).

[6] Damon, W. W., and Schramm, R., "A Simultaneous Decision Model for Production, Marketing and Finance," *Management Science*, **19** 2 (October 1972).

[7] Dorfman, R., and Steiner, P., "Optimal Advertising and Optimal Quality," *American Economic Review* (December 1954).

[8] Eilon, S., "On Smoothing Shipments–A Comment," *Management Science* (May 1971).

[9] Fiacco, A. V., and McCormick, G. P., *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley & Sons (1968).

[10] Hansmann, F., and Hess, S. W., "A Linear Programming Approach to Production and Employment Scheduling," *Management Technology*, No. 1 (1960).

[11] Holt, C., Modigliani, F., Muth, J., and Simon, H., *Planning Production Inventories and Work Force,* Prentice-Hall (1960).
[12] Holt, C., "Dynamic Pricing and Economic Instability–A Comment," *Management Science* (May 1971).
[13] Jaaskelainen, V., "Goal Programming Model of Aggregate Production Planning," *Swedish Journal of Economics* (1969).
[14] Jones, C. H., "Parametric Production Planning," *Management Science,* **15** 11, pp. 843-66 (July 1967).
[14a] Kotler, P., *Marketing Decision Making,* Holt, Rinehart and Winston (1971).
[15] Montgomery, D. B., and Urban, G. L., *Management Science in Marketing,* Prentice-Hall (1969).
[16] Nerlove, M., and Arrow, K. J., "Optimal Advertising Policy under Dynamic Conditions," *Economica* (May 1962).
[17] Ostheimer, R. H., "Frequency Effects Over Time," *Journal of Advertising Research,* **10** 1 (February 1970).
[18] Palda, K., *The Measurement of Cumulative Advertising Effects,* Prentice-Hall (1964).
[19] Peterson, R. F., "Optimal Smoothing of Shipments in Response to Orders," *Management Science* (May 1971).
[20] Powell, M. J. D., "An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivates," *Computer Journal,* **7** (1964).
[21] Quandt, R. E., "Estimating Advertising Effectiveness: Some Pitfalls in Econometric Methods," *Journal of Marketing Research* (May 1964).
[22] Ray, Michael, Sawyer, Alan G., and Strong, Edward C., "Frequency Effects Revisited," *Journal of Advertising Research,* **11** 11, pp. 14-20 (February 1971).
[23] Rao, A. G., *Quantitative Theories in Advertising,* J. Wiley and Sons, New York (1970).
[24] Rao, V., "Alternative Econometric Models of Sales Advertising Relationships," *Journal of Marketing Research* (May 1972).
[25] Sasieni, M. W., "Optimal Advertising Expenditure," *Management Science,* **18** 4 (Part II) (December 1971).
[26] Schneider, E., "Produktion und Lagerhaltung bei einfacher Produktion," *Archiv fur Matematische Wirtschafts und Sozial forschung,* Band IV, Heft 1, Leipzig (1938).
[27] Silver, E. A., "A Tutorial on Production Smoothing and Workforce Balancing," *Operations Research* (Nov.-December 1967).
[28] Stewart, J. B., "Repetitive Advertising in Newspapers: A Study of Two New Products," Graduate School of Business Administration, Harvard University, Boston, Massachusetts (1964).
[29] Simon, J. L., "The Effect of Advertising on Liquor Brand Sales," *Journal of Advertising Research,* **11** 1 (February 1971).
[30] Simon, J. L., *Issues in the Economics of Advertising,* University of Illinois Press, Urbana, Illinois (1970).
[31] Simon, J. L., *The Management of Advertising,* Prentice-Hall, Englewood Cliffs, New Jersey (1971).
[32] Taubert, W. H., "A Computer Search Solution of the Aggregate Scheduling Problem," *Management Science,* **14** 6 (February 1968).
[33] Tuite, M. F., "Merging Marketing Strategy Selection and Production Scheduling: A Higher Order Optimum," *Journal of Industrial Engineering* (February 1968).
[34] Tull, D. S., "The Carry Over Effects of Advertising," *Journal of Marketing* (April 1965).
[35] Vidale, M. L., and Wolfe, H. B., "An Operations Research Study of Sales Response to Advertising," *Operations Research* (June 1957).
[36] Welam, U. P., "Joint Optimization of Production and Marketing Decisions," working paper, School of Management, Boston University.
[37] Welam, U. P., "Multi-Item Production Smoothing Models with almost Closed Form Solutions," *Management Science* (May 1975).
[38] Zangwill, W. I., "Minimizing a Function without Calculating Derivatives," *Computer Journal,* **10** (1967-68).

Professor Ulf Peter Welam is Assistant Professor of Operations Management in the School of Management, Boston University. He received the MBA degree from the Gothenburg School of Economics and Business Administration, and the MS and PhD degrees from Carnegie Mellon University. Currently Dr. Welam is developing a computer system for evaluating maintenance strategies for rail mass transit fleets and is conducting research in aggregate planning models and functional coordination. He is a member of TIMS and the American Accounting Association.

An Efficient Procedure for Implementing a Dual Simplex Network Flow Algorithm

RICHARD V. HELGASON

JEFF L. KENNINGTON
MEMBER, AIIE

Department of Industrial Engineering and Operations Research
Southern Methodist University
Dallas, Texas 75275

Abstract: This paper shows that for linear programming formulations of network flow problems, the nonzero components of rows of the basis inverse are identical. A simple algorithm for identifying these nonzero components is given along with a suggested data structure for implementation. The algorithm requires only one bit of storage for each node plus one additional bit. Finally we indicate how these ideas may be used in the development of a dual simplex code for network flow problems.

■ Minimal cost network flow problems arise whenever items must be shipped through a capacitated network in response to demand. Problems of this type occur in analyses in many areas, including production-distribution systems, urban traffic systems, railways systems, military logistics systems and communication systems (see [8, 9]).

The *minimal cost network flow problem* also called the *capacitated transshipment problem* may be simply stated in terms of a distribution problem in which there are N nodes (plants, warehouses, customers) and M arcs connecting these nodes. It is assumed that an arc $e_m = (i, j)$ may have flow from node i to node j but not conversely. If it is possible to also have shipments from node j to node i, then there must be another arc, say e_k, with $e_k = (j, i)$. Arcs represent both shipping routes currently in use and potential shipping routes. Each node n has associated with it a requirement r_n. If $r_n > 0$, then node n is a supply point with supply equal r_n. If $r_n < 0$, then node n is a demand point with demand equal $-r_n$. Nodes with $r_n = 0$ are called transshipment points. Typically, regional warehouses are modelled as transshipment nodes. Units may be shipped across any arc, $e_m = (i, j)$, at a shipping cost per unit of c_m. Further, each arc e_m has capacity u_m. The objective is to determine which arcs to use and the shipment size so that the total transportation cost of meeting demand, given supply and capacity constraints, is minimal.

A mathematical description of the minimal cost network flow problem is now presented. Let V denote the set of nodes numbered 1 through N and E denote the set of arcs. Let $A_n \subset E$ denote the set of arcs of the form $e_m = (n, j)$, (i.e. arcs after node n); and $B_n \subset E$ denote the set of arcs of the form $e_m = (i, n)$, (i.e. arcs before node n). Then the minimal cost network flow problem may be stated in terms of a highly structured linear program as follows:

$$\min \sum_{e_m \in E} c_m x_m \tag{1}$$

$$\text{st.} \sum_{e_m \in A_n} x_m - \sum_{e_m \in B_n} x_m = r_n \text{ for all } n \in V \tag{2}$$

$$0 \leqslant x_m \leqslant u_m, \text{ for all } e_m \in E, \tag{3}$$

where it is assumed that $\Sigma_{n \in V} r_n = 0$. If for some problem $\Sigma_{n \in V} r_n < 0$, then there is no feasible solution. If $\Sigma_{n \in V} r_n > 0$, then the problem can be converted to the prescribed form by adding a dummy destination with a requirement of $-\Sigma_{n \in V} r_n$ and slack arcs from each source to the dummy destination. These slack arcs are given unit costs of zero and infinite arc capacities.

Received April 1976; revised August 1976.

Many studies concerning solution techniques for distribution problems frequently involve subprograms which are minimal cost network flow problems. Solution of the distribution problem frequently requires that the subprograms be solved successively with one or more bound changes, one or more cost coefficient changes, or both. In most cases, a LP basis for the subprogram is available for which the optimality criteria is satisfied, but one or more variables exceeds its bound. This precise situation arises in the following studies, [2, 13-15, 17]. When this situation occurs one may wish to apply a dual simplex method to regain feasibility for these subprograms. Although much work has appeared in the literature concerning efficient implementations of the primal simplex method for network problems, very little has appeared concerning the dual simplex method, (see [3, 5, 10-12, 15, 16, 18]). The single exception is a paper by Glover, Karney, and Klingman [4]. We believe that this is primarily due to the fact that the techniques are quite similar computationally and an efficient implementation of the primal simplex method should lead to an equally good implementation of the dual simplex method. The only difference in the methods, from a computational point of view, is that the dual method requires all the computational machinery of a primal method plus the capability of generating rows of the current basis inverse. These rows are required for performing dual simplex pivots but are not required for primal simplex pivots. This is due to the fact that the rule for selecting the entering column using the dual simplex method involves a ratio of the reduced cost and an updated row of the tableau (i.e. a row of the basis inverse multiplied by the nonbasic columns). The rule for selecting the entering column using the primal simplex method involves only the reduced cost. Hence, a good implementation of the dual simplex method for (1) - (3) requires an efficient means for generating rows of the inverse of the current working basis.

This paper shows that there are at most two distinct elements in any row of the inverse of a working basis for the minimal cost network flow problem. Furthermore, if there are two distinct elements in a row, then one of these must be zero. This result and the constructive proof presented leads to an extremely simple algorithm for generating rows of this inverse. The algorithm requires only $N+1$ bits of storage over and above that required for a primal simplex procedure. This is in contrast to the dual code of Glover, Karney, and Klingman [4] which uses three additional node length arrays. We are not suggesting that the dual technique be used to solve the minimal cost network flow problem from a "dead" start. Evidence has been presented (see [4, 7]) that dual codes are slower than primal codes for solving problems for which no prior information is available. Even though the aforementioned studies did not include the new results presented in this paper, we suspect that the conclusion would still hold for large network problems. We believe that the potential value for the results presented in this paper lies in those situations in which an extremely good dual starting solution is available.

We now present the notations and conventions used in this exposition. Lower case Greek and Latin letters underlined denote column vectors. $\underline{\beta}'$ denotes the transposition of the column vector $\underline{\beta}$ while β_i denotes the i*th* component of $\underline{\beta}$. A vector of all ones is denoted by $\underline{1}$. The vector $\underline{e}_i$ has a one in position i and zeroes elsewhere.

Basis Property for Network Problems

In this section we present the known results concerning the graphical interpretation of bases for the minimal cost network flow problem. The graph associated with (1) - (3) is defined by the node set V and the arc set E. A *simple path* in $[V, E]$ is a sequence of alternating nodes and connecting edges such that no node (or arc) is repeated. The direction of the arcs is unimportant in defining a simple path and both $i,(i,j),j,(j,k),k$ and $i,(j,i),j,(j,k),k$ are simple paths connecting nodes i and k. A *cycle* is a simple path together with an arc joining the beginning node and ending node of the path. A *connected graph* is a collection of nodes and arcs having at least one simple path between every pair of nodes, and a *tree* is a connected graph with no cycles.

Since the constraints (2) have rank equal $N-1$ (see [1]), we add an artificial variable to some row (node), say row ℓ, to obtain the following system.

$$\left.\begin{aligned} \sum_{e_m \in A_\ell} x_m - \sum_{e_m \in B_\ell} x_m + a &= r_\ell \\ \sum_{e_m \in A_n} x_m - \sum_{e_m \in B_n} x_m &= r_n, \text{ for all } n \in V, \; n \neq \ell \end{aligned}\right\} . \quad (4)$$

The above system has rank N and we assign zero to both the cost coefficient and upper bound of a. Let B be any basis of (4). Clearly a appears in every basis, B. Then the subgraph generated by B, called F_B, consists of all vertices of V and arcs corresponding to columns of (4) in B. The artificial variable may be thought of as an arc incident to a single node, i.e. node ℓ. It is well known that F_B is a tree, if and only if B is a basis. Node ℓ is called the root of the tree. Further, by a proper permutation of rows and columns, B can be placed in lower triangular form; and hence B^{-1} can be placed in upper triangular form. A proof of these results can be found in [1].

To make these remarks concrete a simple example is presented. Consider the network shown in Fig. 1 and let node 3 be the root. The constraints corresponding to (4) are simply,

x_1	x_2	x_3	x_4	x_5	x_6	a		
1	1							10
-1		-1	1	1			=	0
	-1	1	-1		1	1		0
				-1	-1			-10

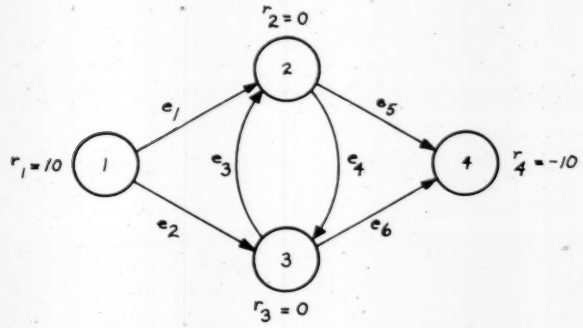


Fig. 1. Example of a minimal cost network flow problem.

A basis consisting of the columns corresponding to x_1, x_4, x_5, and a results in the tree shown in Fig. 2.

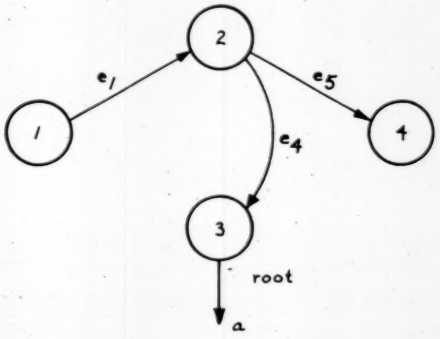


Fig. 2. F_B for basis whose columns are x_1, x_4, x_5, and a.

Basis Inverse Property for Network Problems

In this section we present a new property exhibited by inverses of working bases for minimal cost network flow problems. Given this property, we can readily construct rows of this inverse and easily execute dual simplex pivots.

Proposition. Let B be any working basis for the minimal cost network flow problem as described by constraints (4). Then the nonzero components of any row of B^{-1} are identical and are either 1 or -1.

Proof. Let $\underline{\beta}'$ denote the t*th* row of B^{-1}. Then

$$\underline{\beta}' B = \underline{e}_t' . \tag{5}$$

Denote the arcs corresponding to columns of B, excluding the artificial arc, by $e_{m_1} = (i_1, j_1)$, $e_{m_2} = (i_2, j_2), \ldots, e_{m_p} = (i_p, j_p)$.

Case 1. Suppose arc $e_{m_k} = (i_k, j_k)$ appears in the t*th* column of B. From (5) we have

$$\beta_\ell = 0, \text{ and} \tag{6}$$

$$\beta_i - \beta_j = \left\{ \begin{array}{l} 1, \text{ if } i = i_k \text{ and } j = j_k \\ 0, \text{ otherwise.} \end{array} \right\} . \tag{7}$$

Rewriting (7) we have,

$$\beta_{i_k} - \beta_{j_k} = 1, \text{ and } \beta_i = \beta_j \text{ for } i \neq i_k \text{ and } j \neq j_k . \tag{8}$$

Since B is a basis, the corresponding tree F_B, will have one of two forms as shown in Fig. 3, with all nodes not shown grouped together in appropriate rectangles. This gives rise to two subcases 1a and 1b, depending upon the orientation of

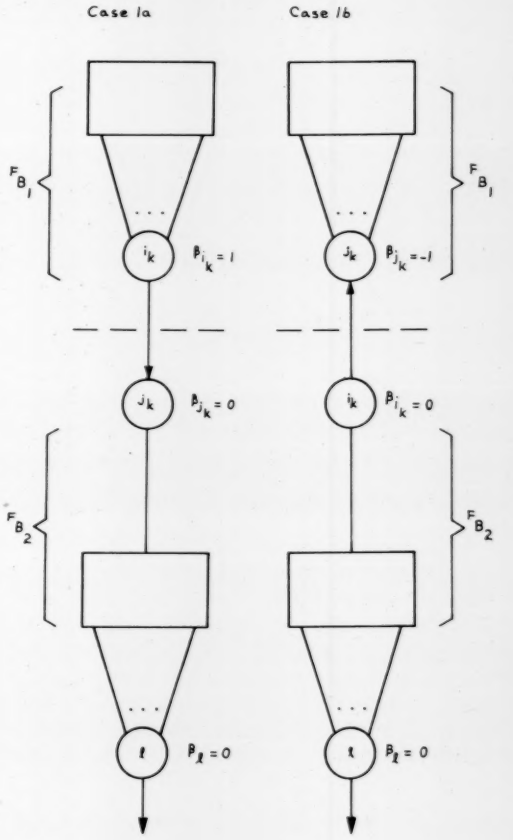


Fig. 3. Basis trees for the two cases of F_B

the arc (i_k, j_k). If the arc (i_k, j_k) is removed from F_B, two trees, F_{B_1} and F_{B_2} result. In both trees, (8) must hold for all basic arcs. Hence, all nonzero components of the t*th* row of B^{-1} are identical and are either 1 or - 1.

Case 2. Suppose the artificial variable appears in the t*th* column of B. Then from (5) $\beta_\ell = 1$ and $\beta_i = \beta_j$ for all arcs $e_{m_1}, \ldots, e_{m_p}$. Hence, $\underline{\beta}' = \underline{1}'$. This completes the proof of the proposition. The above proof shows explicitly how one may construct rows of the inverse. For the particular row corresponding to the basic arc (i_k, j_k), one need only identify the tree F_{B_1}.

Data Structure and Algorithm for Generating Rows of Inverse

Before describing the algorithm for generating rows of the basis inverse, we first present a data structure which may be used to define the trees corresponding to the F_B's for different bases. For purposes of explanation, we draw the basis trees with the root at the bottom as shown in Fig. 4.

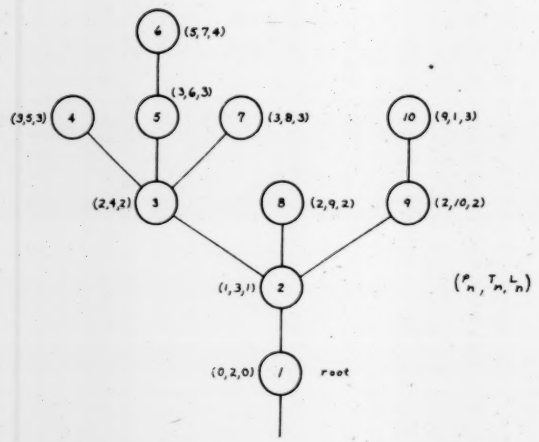


Fig. 4. Example of a tree with labels.

Each node has three labels which are used to describe the tree. The *level* of node n, denoted by L_n, is defined to be the number of arcs in the path from n to the root. The level of the root node is defined to be zero. The *predecessor* of node n, denoted by P_n, is defined to be the node on the path from n to the root whose level is $L_n - 1$. The predecessor of the root node is also defined to be zero. Node m is said to lie above node n if n lies on the unique path from m to the root. Let $U_n \subset V$ denote the set of nodes above n. Given any tree, it is possible to order the nodes so that for every node n, the set U_n follows immediately after n. For the tree of Fig. 2, two such orderings are $\{3, 2, 1, 4\}$ and $\{3, 2, 4, 1\}$. Given an ordering with the prescribed property, the *thread* of node n, denoted by T_n, indicates the node following n in that ordering. We have defined the thread of the last node in the ordering to be the root. The thread and level labels together allow one to directly identify F_{B_1} (see Fig. 3).

We present an algorithm for determining a set of thread labels for any tree. Let ℓ denote the root of the tree.

(Alg-1) Algorithm to Determine Thread Labels

0. [Begin at root] $i \leftarrow 1$ and $t_1 \leftarrow \ell$
1. [Find candidates in subtree above t_i] $\Gamma_{t_i} = \{k: P_k = t_i\}$. If $\Gamma_{t_i} \neq \Phi$, let t_{i+1} be any member of Γ_{t_i} and go to 4.
2. [No subtree above t_i, find new candidates for next ordering position] Let $\Omega_{t_i} = \{\ell\} \cup \{r: r \neq t_j$ for all $j \leqslant i$ and P_r lies on the path from t_i to the root$\}$.
3. [Select a candidate at highest level] Let t_{i+1} be such that $L_{t_{i+1}} = \underset{r \in \Omega_{t_i}}{\text{MAX}} \; L_r$.
4. [Assign thread] $T_{t_i} \leftarrow t_{i+1}$.
5. [Update or terminate] If $t_{i+1} = \ell$, stop; otherwise $i \leftarrow i+1$ and go to 1.

The above procedure can be used to obtain the threads given only the predecessor and the level labels. In particular, ALG-1 could be applied to any starting basis tree to obtain a complete set of thread labels. Figure 4 gives a tree and a corresponding set of labels with the threads determined by the above algorithm.

This particular data structure is a combination of ideas presented in three different papers. Johnson [11] suggested using three labels for describing trees one of which was the predecessor. Glover, Klingman, and Stutz [6] improved on Johnson's data structure by replacing two of Johnson's labels with the thread. More recently Bradley and Brown [3] suggested the use of the level with a different data structure. We believe that the combination of the predecessor, thread, and level provides one of the most powerful three label data structures available for implementing both primal and dual, single and multicommodity network algorithms.

We now present an algorithm for determining a row of the inverse of any basis from the system (4). Suppose we wish to determine the t*th* row of B^{-1} and the t*th* column of B corresponds to $e_{m_k} = (i_k, j_k)$. Define n as follows:

$$n = \begin{cases} i_k, \text{ if } L_{i_k} > L_{j_k} \\ j_k, \text{ otherwise.} \end{cases}$$

If the t*th* column of B corresponds to the artificial variable a, set n equal to the root. The following algorithm determines the nonzero components of the t*th* row of B^{-1}.

(ALG-2) Algorithm to Determine Nonzero Elements of a Row of B^{-1}

1. [Row for artificial is 1'] If n = root, $\beta_j^* \leftarrow 1$ for all j and terminate.
2. [Initialize row] $\beta_j^* \leftarrow 0$ for all j.
3. [Save level for input node] $L \leftarrow L_n$.
4. [Begin tree search at n] $j \leftarrow n$.
5. [Position j is nonzero] $\beta_j^* \leftarrow 1$.
6. [Follow thread] $j \leftarrow T_j$.
7. [U_n complete?] $L_j > L$, go to 5; otherwise terminate.

The first step of the above algorithm is justified by case 2 of the proof of the proposition. That is, if the t*th* column of B is the artificial variable, then the t*th* row of B^{-1} is $\underline{1}'$. The remaining steps are justified since the procedure has simply identified F_{B_1} (see Fig. 3) in case 1 of the proof of the proposition and applied (8) to this tree. Note that the nodes of F_{B_1} are given by U_{j_k} in case 1a and by U_{i_k} in case 1b. For the example of Fig. 4 with $n = 3$, the nonzero β's are those associated with nodes 3 through 7.

Since the algorithm determines only the nonzero components of a row of the inverse, the storage required is one bit per node. One additional bit is required to indicate whether the case 1a or 1b holds. It should be noted that in a dual simplex code, the first step of ALG-2 is not needed since the artificial variable, a, cannot leave the basis. Furthermore, step 2 need not be executed at each pivot. That is, after the row of the inverse has been used, the nonzero components could be replaced by zero using a similar scheme.

Implementation of Dual Simplex Method

This section indicates how ALG-2 may be used in the development of a specialized dual simplex code for the minimal cost network flow problem. Once the entering and leaving variables have been selected, dual and primal pivots are identical. Since the leaving variable for the dual method is any one which violates a constraint of the form $0 \leqslant x_j \leqslant u_j$, we concentrate only on the rule for selecting the entering variable in a simple upper-bounded context, wherein the constraints (3) do not appear explicitly in any basis.

Let $I_0 = \{j: x_j$ nonbasic and $x_j = 0\}$ and $I_u = \{j: x_j$ nonbasic and $x_j = u_j\}$. Suppose the leaving variable, say x_k, appears in the i*th* column of B. Then the entering variable, x_{j_0} is selected by the following rules,

Case I. $(x_k < 0)$

$$\frac{\bar{c}_{j_0}}{y_{ij_0}} = \text{MIN}\left\{ \underset{j\in I_0}{\text{MIN}}\left[\frac{\bar{c}_j}{y_{ij}} : y_{ij} < 0\right], \underset{j\in I_u}{\text{MIN}}\left[\frac{\bar{c}_j}{y_{ij}} : y_{ij} > 0\right]\right\}$$

Case II. $(x_k > u_k)$

$$\frac{\bar{c}_{j_0}}{y_{ij_0}} = \text{MAX}\left\{ \underset{j\in I_0}{\text{MAX}}\left[\frac{\bar{c}_j}{y_{ij}} : y_{ij} > 0\right], \underset{j\in I_u}{\text{MAX}}\left[\frac{\bar{c}_j}{y_{ij}} : y_{ij} < 0\right]\right\}$$

where $\bar{c}_j$ is the reduced cost for x_j, and y_{ij} is the i*th* row of the basis inverse multiplied by the original column in (4) corresponding to x_j. The y_{ij}'s are also commonly referred to as components of an updated row. In terms of our previous notation,

$$y_{ij} = \underline{\beta}'_i \underline{a}_j$$

where $\underline{a}_j$ denotes the column of (4) corresponding to x_j and $|\underline{\beta}_i| = \underline{\beta}^*$, as obtained from ALG-2. Suppose x_j corresponds to arc $e_j = (r, s)$. Then,

$$\underline{a}'_j = [0, \ldots, 0, \underset{\uparrow}{1}, 0, \ldots, 0, \underset{\uparrow}{-1}, 0, \ldots, 0]$$
$$r\text{th position} \quad s\text{th position}$$

Hence,

$$y_{ij} = \begin{cases} \beta_r - \beta_s, \text{ if case 1a occurs,} \\ \beta_r + \beta_s, \text{ if case 1b occurs.} \end{cases}$$

Since the β's are restricted to values from exactly one of the sets $\{-1,0\}$ or $\{1,0\}$, it follows that $y_{ij} \in \{-1,0,1\}$. Consequently, the rules for the entering variable can be further simplified. Note that if Case II occurs, x_k must be set to its upper bound, u_k.

Summary and Conclusions

This exposition presents new results which, we believe, will lead to a faster dual simplex code for network flow problems than those developed and tested to date. Glover, Karney, and Klingman [7] indicate that a large percentage of the computational time for their dual code involves selecting the entering variable. The ideas presented in this paper are directed at reducing the computational burden of precisely this part of a dual simplex algorithm. We assume that an implementation of these ideas would use the efficient computational machinery available for executing pivots in network codes once the entering and leaving variables have been selected. Although, implementation of these ideas will make the dual simplex method more competitive with a primal method in solving problems from a "dead" start; we believe, that the major use of these results should be for situations in which good starting dual feasible bases are available.

References

[1] Bazaraa, M. S., and Jarvis, J. J., *Linear Programming and Network Flows*, Published by Georgia Institute of Technology, Atlanta, Ga. (1974).

[2] Barr, Richard, Glover, Fred, and Klingman, Darwin, "Development and Computational Testing of a Fixed Charge Transportation Code," Presented at the Joint National Meeting of ORSA and TIMS in Puerto Rico (October 1974).

[3] Bradley, Gordon H., and Brown, Gerald G., "A Comparison of Storage Structures for Primal Network Codes," Presented at the Joint National Meeting of ORSA and TIMS in Chicago, (April 1975).

[4] Glover, Fred, Karney, D., and Klingman, D., "Double-Pricing Dual and Feasible Start Algorithms for the Capacitated Transportation (Distribution) Problem," Research Report 105, Center for Cybernetic Studies, The University of Texas, Austin, Texas (May 1970).

[5] Glover, Fred, Karney, D., Klingman D., and Napier, A., "A Computational Study on Start Procedures, Basis Change Criteria, and Solution Algorithms for Transportation Problems," *Management Science*, **20**, 5, 793-813 (January 1974).

[6] Glover, F., Klingman, D., and Stutz, J., "Augmented Threaded Index Method for Network Optimization," *INFOR*, **12**, 3, 293-298 (October 1974).

[7] Glover, F., Karney, D., and Klingman, D., "Implementation and Computational Comparisons of Primal, Dual, and Primal-Dual Computer Codes for Minimum Cost Network Flow Problems," *Networks*, **4**, 3, 191-212 (July 1974).

[8] Glover, Fred, and Klingman, Darwin, "Real World Applications of Network Related Problems and Breakthroughs in Solving Them Efficiently," *ACM Transactions on Mathematical Software*, **1**, 1 47-55 (March 1975).

[9] Glover, Fred, and Klingman, Darwin, "Network Application in Industry and Government," Research Report 247, Center for Cybernetic Studies, The University of Texas, Austin, Texas (September 1975).

[10] Johnson, Ellis, "Programming in Networks and Graphs," ORC Report 65-1, University of California–Berkeley (January 1965).

[11] Johnson, Ellis, "Networks and Basic Solutions," *Operations Research*, **14**, 4, 619-623 (July-August 1966).

[12] Karney, D., and Klingman, D., "Implementation and Computational Study of an In-Core Out-Of-Core Primal Network Code," Research Report 158, Center for Cybernetic Studies, The University of Texas, Austin, Texas (October 1973).

[13] Kennington, Jeff and Shalaby, Mohamed, "An Effective Subgradient Procedure for Minimal Cost Multicommodity Flow Problems, (Forthcoming in *Management Science*, 1976).

[14] Kennington, Jeff and Unger, Ed., "A New Branch-And-Bound Algorithm for the Fixed-Charge Transportation Problem," *Management Science* **22**, 10, 1116-1126, (June 1976).

[15] Langley, Robert W., "Continuous and Integer Generalized Flow Problems," (Unpublished Dissertation), Dept. of Industrial and Systems Engineering, Georgia Tech. (June 1973).

[16] Langley, Robert W., Kennington, Jeff, and Shetty, C. M., "Efficient Computational Devices for the Capacitated Transportation Problem," *Naval Research Logistics Quarterly*, **21**, 4, 637-647 (December 1974).

[17] Rardin, Ron L., "Group Theoretic and Related Approaches to Fixed Charge Problems," (Unpublished Dissertation), Dept. of Industrial and Systems Engineering, Georgia Tech. (December 1973).

[18] Srinivasan, V., and Thompson, G. L., "Benefit-Cost Analysis of Coding Techniques for the Primal Transportation Algorithm *Journal of the ACM*, **20**, 2, 194-213 (April 1973).

Mr. Richard V. Helgason is an Operations Research Analyst with Texas Instruments, Inc. and a graduate student in the Department of Industrial Engineering and Operations Research at Southern Methodist University. He received a BS in Mathematics from Iowa State University, an MS in mathematics from the University of Wisconsin, and an MS in Engineering Design and Economic Evaluation from the University of Colorado. He is a member of ACM, ORSA, SIAM, and the Mathematical Programming Society.

Dr. Jeff Kennington is an Assistant Professor in the Department of Industrial Engineering and Operations Research at Southern Methodist University in Dallas, Texas. He received his PhD from the Department of Industrial and Systems Engineering, Georgia Institute of Technology in Atlanta, Georgia. Research interests include integer programming and network flow theory. His other publications have appeared in *Operations Research, Management Science, Naval Research Logistics Quarterly*, and *AIIE Transactions*. Professor Kennington is a member of AIIE, ORSA, TIMS, and the Mathematical Programming Society.

Scheduling Multispecialty Clinics Using Integer Programming

THOM J. HODGSON, P.E.
SENIOR MEMBER, AIIE
Department of Industrial and Systems Engineering

KERRY E. KILPATRICK
SENIOR MEMBER, AIIE
Health Systems Research Division
J. Hillis Miller Health Center

IRA M. LONGINI
ASSOCIATE MEMBER, AIIE
Health Systems Research Division
J. Hillis Miller Health Center

The University of Florida
Gainesville, Florida 32611

Abstract: This paper describes an integer programming approach to scheduling multispecialty outpatient clinics in a university health center. Using maximum expected cross-consultations as an objective, approximately thirty clinics have been scheduled for a five-day cycle. Each clinic has its own medical staff but shares space, personnel, and ancillary services with other clinics on a rotating basis. If two specialties requiring cross-consultation do not meet at the same time, the consultation process is impaired. The clinics must also be scheduled such that the available capacities of the common resources are not exceeded. In addition, the scheduling process must consider other constraints such as the commitments of the physicians to teaching, hospital rounds, and operating room schedules. This problem has been formulated and solved as a 0-1 integer quadratic program to maximize the expected cross-consultation volume subject to constraints on shared resources. The paper discusses the process by which required input data were collected and the results of the implementation of the model.

■ The J. Hillis Miller Health Center, similar to many medical centers, operates a multispecialty outpatient clinic system. The annual patient volume through the clinics is approximately 130,000 patient visits. The various medical specialties are assigned meeting times and space on a repetitive five day cycle. Some specialties such as opthalmology and pediatrics have permanently assigned space because of special requirements. Approximately thirty other clinics share space, ancillary services, allied health personnel and other common resources on a rotating basis.

Finding the best assignment of the number of times each of these thirty specialties will hold a clinic each week, the number of rooms at each meeting, and the meeting times, constitutes the scheduling problem. Each specialty requires a specified number of room-hours each week to serve its outpatient volume adequately. This requirement places a lower bound on the room-hours assigned to the specialty. The common resources also constrain the schedule; personnel such as nurses are limited and the capacity of radiology and laboratories to serve the clinics cannot be exceeded. The attending staff also have preferred operating room schedules, times when conferences and rounds must be held, and (in a teaching environment) times for scheduled lectures, laboratories, seminars, and research activities. These factors further constrain the schedule by limiting the number of feasible meeting time-room assignment combinations.

Received January 1975; revised July 1976.

Problem Statement

When the study began, each specialty clinic met its patients essentially independently of all other clinics. Patients were scheduled independently, lab and radiology reports were sent to the individual attending physicial or resident, and professional and clinic charges were captured when they were generated, but billed centrally. In preliminary discussions of the scheduling problem with the attending physicians, a vague notion arose of a physician's "preference" for having his clinic scheduled at the same time as another specialty. This preference included the subjective probability of cross-consultations between physicians in different clinics and the subjective probability that a patient

This research is supported in part by ONR Contract No. N00014-68-A-0173-0013 and by the Health Systems Research Division, University of Florida.

scheduled for one specialty may also be scheduled for another specialty during the same visit. In both cases, proper scheduling of the clinics would benefit the patient by minimizing the number of days the patient would have to visit the clinics. Although the physician orders the consult and consults occur between physicians, the patient must be present at the time of the consult for it to be considered as valid in the model.

An early attempt to quantify these preferences through a relationship chart approach met with limited success. We then decided to quantify the actual historical clinic interaction by counting cross-consultations that had occurred between clinic specialties. This approach also proved difficult because cross-consultations were buried in patient records. Further, what really was required was not the number of referrals that *had* occurred, but that *would* occur if two clinics met concurrently. To obtain these data we requested each of the thirty attending physicians consider their relationship with the remaining twenty-nine clinics and to respond to the following question: "If your clinic met at the same day and time as specialty X, for what percentage of your patients would you order cross-consultations with X?" These values, multiplied by the referring clinic's patient volume, provide a measure of the potential one-way cross-consultation flows. Additional insights into the problems and benefits of this data collection process will be discussed in a later section.

The formulation of the model, in terms amenable to solution, evolved through numerous meetings with physicians, administrators and other staff personnel. The physicians felt that the anticipated volume of cross-consultations should be the primary measure of clinic interaction. Further, to explicitly represent the patients' requirements, it was decided that the model should only consider a cross-consult valid when the patient was physically present on the day the consultant's clinic met. Thus, merely passing lab results or conferring over a radiographic film was not considered to be a cross-consult in the model.

To illustrate the use of the cross-consult measure of clinic interaction, consider two hypothetical clinics: Clinic I meets three days a week and needs to consult with Clinic II on 10% of its patients. Clinic II meets two days a week and needs to consult with Clinic I on 15% of its patients. Clinic I has a patient volume of 20 patients per day, and Clinic II has 30. The expected number of consultations in both directions would be $0.10 \times 20 = 2$, and $0.15 \times 30 = 4.5$ patients per day respectively. The two-directional flow is 6.5 patients per day, given that the clinics meet concurrently. Consider schedule C in Table 1. If, on Monday, it was necessary for a patient to have a consultation with Clinic II, the patient would have to return on Tuesday for the consultation. Then he would have to return on Wednesday in order to complete treatment with Clinic I. This situation is inefficient and inconvenient for both the patient and the medical staff. If, however, schedules A or B were in effect, both the consultation and the treatment could have been completed on Monday. Since schedule B has two days in which the schedules for Clinics I and II coincide, it is the most desirable. As noted, for purposes of modeling the scheduling process, it is assumed that there can be no cross-consults between clinics which do not operate concurrently. This concept can also be applied to clinics operating only in the morning or only in the afternoon. If a clinic meets only in the morning and another meets only in the afternoon, there can be no cross-consults between them.

Table 1: Sample cross-consult calculation.

Schedule	Clinic II	Days on which clinic meets M	T	W	Th	F	Expected Level of cross-consultation per week.
A	Clinic I	x		x		x	6.5
	Clinic II	x	x				
B	Clinic I	x		x		x	13
	Clinic II	x		x			
C	Clinic I	x		x		x	0
	Clinic II		x		x		

The objective, then, of the clinic scheduling model is to maximize the sum of the expected cross-consults between all clinics. It could be argued that some consultations are more important than others. However, for the case in point, there was general agreement among both the medical and administrative staffs of the health center that the objective was fair to all concerned.

If all clinic specialties could meet all day, every day, cross-consults would be maximum. Clearly, this is infeasible for several reasons. First, there is a minimum number of rooms which each specialty must have per clinic; otherwise physician manpower cannot be used efficiently. Also, the other demands on physician time such as operating schedules, teaching and research commitments preclude certain meeting times. This, fortunately reduces the number of feasible solutions that must be considered.

As noted earlier, the problem is further constrained by the capacity of ancillary services and the number of support personnel, both of which are fixed in the short run. If certain clinics are scheduled concurrently (even though they have a potentially high number of cross-consults) the resulting loads on ancillary services may exceed the capacity. For example, both the general medical clinic and urology are relatively high users of clinical laboratories. They also experience a high number of cross-consults. The solution procedure must, therefore, attempt to have as many of the scheduled meeting times of these two clinics be concurrent as possible without exceeding laboratory capacity. The objective function and constraints are formulated algebraically in the next section.

Formulation

The problem was formulated as a 0-1 integer program with a quadratic objective function. To present the basic formulation, the following definitions are necessary:

X_{ij} = 1, if schedule j for clinic i is chosen
= 0, otherwise,

N = Total number of clinics in the system,

A_{ijk} = The number of rooms clinic i requires in schedule j during period k,

M_k = The total number of rooms available for scheduling period k,

$I_{i\ell}$ = The expected number of patients per period requiring consultation between clinic i and ℓ given that they meet concurrently,

$P_{i\ell jm}$ = The number of periods that clinics i and ℓ meet concurrently in a week, given schedules j and m are chosen,

$n(i)$ = The number of possible schedules to be considered for clinic i,

B_{ijks} = The average load generated in scheduling period k by clinic i, schedule j, on ancillary service s (in work units),

C_{ks} = Capacity of ancillary service s (in work units) in period k and

S = Number of ancillary services considered.

Note that $P_{i\ell jm}I_{i\ell}$ is the expected number of cross-consults that would occur between clinics i and ℓ if they were to meet concurrently.

Because $X_{ij}X_{\ell m} = 1$ if schedule j is chosen for clinic i and schedule m is chosen for clinic ℓ and $X_{ij}X_{\ell m}$ is zero otherwise, the objective function can be represented algebraically as:

$$\text{Maximize} \sum_{i=1}^{N} \sum_{\ell=1}^{N} \sum_{j=1}^{n(i)} \sum_{m=1}^{n(\ell)} P_{i\ell jm} \cdot I_{i\ell} \cdot X_{ij} \cdot X_{\ell m} \quad (1)$$

Since one and only one decision must be made for each clinic, the following constraint set is necessary:

$$\sum_{j=1}^{n(i)} X_{ij} = 1, i = 1, 2, \ldots, N. \quad (2)$$

The room availability constraints are as follows:

$$\sum_{i=1}^{N} \sum_{j=1}^{n(i)} A_{ijk}X_{ij} \leq M_k, k = 1, \ldots, 10 \quad (3)$$

Each day of the week is divided into two scheduling periods (morning and afternoon). Consequently, the subscript k (referring to scheduling periods) in (3) has values ranging from 1 to 10.

The ancillary services constraint can be expressed as follows:

$$\sum_{i=1}^{N} \sum_{j=1}^{n(i)} B_{ijks}X_{ij} \leq C_{ks}, \text{ for all } k, s = 1, \ldots, S \quad (4)$$

Constraints imposed by operating room schedules and teaching schedules can be handled simply by not considering clinic schedules in the models which conflict. This approach reduces the number of decision variables in the model and facilitates obtaining solutions.

An Example Problem

In order to provide additional insight into the model, the following example is given. Consider a three clinic problem with only a two day week. Also, assume that each clinic meets on only one day per week for the full day. For simplicity, let the only constraints be those dealing with room-hours (i.e., no ancillary services). The descriptive data for the problem appear in Table 2, and the 0-1 constraint set for the problem appears (in tabular form) in Table 3.

Table 2: Room – hour requirements.

i	Clinic	A_{i11}	A_{i12}	A_{i21}	A_{i22}
1	Clinic I	8	0	0	8
2	Clinic II	5	0	0	5
3	Clinic III	6	0	0	6

Total Rooms Available = 14

Inter - Clinic Information FLow Matrix

I =

i \ ℓ	1	2	3
1	0	12	5
2	8	0	8
3	5	12	0

Weekly concurrent meeting matrix

P =

	m	1	2	1	2	1	2
	ℓ	1	1	2	2	3	3
j	i						
1	1	0	0	1	0	1	0
2	1	0	0	0	1	0	1
1	2	1	0	0	0	1	0
2	2	0	1	0	0	0	1
1	3	1	0	1	0	0	0
2	3	0	1	0	1	0	0

An optimal solution (one of two) for the problem is:

$$X_{11} = X_{21} = X_{32} = 1, X_{12} = X_{22} = X_{31} = 0$$

The solution indicates that Clinics I and II should be scheduled on day one and Clinic III should be scheduled on day two. The value of the objective function is 12.

Table 3: Constraint set for sample problem.

Clinic I		Clinic II		Clinic III			
X_{11}	X_{12}	X_{21}	X_{22}	X_{31}	X_{32}	(Decision Variables)	
1	1					= 1	
		1	1			= 1	
				1	1	= 1	
8		5		6		⩽ 14	
	8		5		6	⩽ 14	

Quadratic Objective Matrix ($P_{i\ell jm}$ $I_{i\ell}$)

	m	1	2	1	2	1	2
j	i \ ℓ	1	1	2	2	3	3
1	1	0	0	12	0	5	0
2	1	0	0	0	12	0	5
1	2	8	0	0	0	8	0
2	2	0	8	0	0	0	0
1	3	5	0	12	0	0	0
2	3	0	5	0	12	0	0

Data Requirements

The collection and analysis of input data for the model described in previous sections serves to illuminate the practical problems associated with implementation of the model solution in the real environment.

In any organized system where working space is at a premium, there will be competition for space among the various organizational entities within the system. For instance, this phenomenon can certainly be observed between departments in a corporate staff activity, or between departments in a college or university. The provision of adequate space results in better working conditions and more flexibility of operation. Space is power.

In the present case, the allocation of treatment rooms to specialties was a result of various forces, not all of which were directly related to need. Consequently, a work sampling analysis of the utilization of the space assigned revealed considerable variation in apparent utilization among specialties. While these data were not received enthusiastically by the medical staff of underutilized clinics, they did serve to focus the attention of the medical staff on the problem of clinic scheduling. In addition, the work sampling measurements, augmented by direct observations, were used to determine the average processing time per patient by clinic specialty. These processing times, together with the expected visit volume by specialty, determined the number of rooms-hours which were allowed to be assigned to each clinic. The allocations, A_{ijk}, were chosen such that the total room-hours did not exceed the allowable limit in any schedule acceptable to the physicians.

The collection of actual clinic interaction data appeared to be a formidable task. With 30 clinics to schedule, it would have been necessary to gather data on up to 870 separate referral streams. Thus, direct data collection was judged to be beyond the means available to the project. In addition, there was concern that the data would be biased by the present clinic schedule (i.e., if referrals would be inconvenient due to non-overlapping clinic schedules, some referrals might be repressed, biasing the data). As noted above, the attending physicians of each clinic specialty were asked to estimate the percentage of their patients for whom cross-consults would be ordered from the other specialties *if they met concurrently*. This percentage, multiplied by the clinic volume, gave an estimate of the cross-consult flow. This procedure had the following advantages:

1. It was relatively cheap;
2. It was unbiased with respect to the existing schedule;
3. It allowed the physicians to become involved in an administrative procedure of concern to them, thereby facilitating implementation of the new schedule; and
4. It permitted spot checks with actual data for accuracy.

To gather data on possible individual clinic meeting times, operating room schedules and teaching schedules were reviewed to ascertain alternate times during which each clinic specialty might be scheduled. The attending physician of each clinic specialty reviewed the resulting set of alternative clinic schedules to eliminate schedules which violated other miscellaneous constraints. The remaining set of alternative clinic schedules was used as input to the model.

This data collection effort generated inputs for the model which were judged to be adequately accurate. Perhaps of even greater importance, the medical staff was sufficiently involved in the project through the data collection process that they began to anticipate changes in the clinic schedule, thus paving the way for eventual acceptance.

Solution Technique

The quadratic program formulated here is, in general, difficult to solve. The quadratic objective matrices possess little useable structure. Consequently, we used implicit enumeration to obtain optimal solutions. The basic strategies of integer quadratic programming [7, 8, 10] and implicit enumeration [1, 2, 4, 9] are well established. Therefore, we do not restate these strategies and techniques here.

Ferreira and Hodgson [3] have described and documented the algorithm for a similarly structured linear integer 0-1 program. The algorithm is a tree enumeration procedure (one level of the tree for each clinic) which uses the room availability constraints (Eq. (3)), the ancillary services constraints (Eq. (4)), and a bounding function to eliminate (implicitly) large portions (hopefully) of the tree. The only modification to the Ferreira-Hodgson algorithm is the substitution of the quadratic objective: the modification is straightforward and is not discussed here.

The bounding function, while also straightforward, deserves comment. Assume that the algorithm has fathomed the first k levels of the enumeration tree (i.e., the decision variables for the first k clinics have been fixed and the N-k

remaining clinics have not). Also assume that an upper bound (UB) (the best previously obtained solution) for the problem is known, and that the room availability and ancillary service constraints have not been violated. The question, then, is:

> Is the partial objective function corresponding to the value calculated from the fixed decision variables, plus the best possible value obtained from the remaining unfixed variables, greater than the upper bound?

Or symbolically,

$$Q(k) + \phi(k) > UB, \tag{5}$$

where $Q(k)$ is the partial objective function value of the mixed decision variables associated with the first k clinics, and $\phi(k)$ is the best possible value that can be obtained from the remaining decision variables. If the inequality is not satisfied, then it is useless to pursue the current branch of the tree further. $\phi(k)$ is defined as follows:

$$\phi(k) = \sum_{\ell = k+1}^{N} \sum_{i=1}^{N} \max_{j,m} (P_{ij\ell m} \cdot I_{i\ell}). \tag{6}$$

Clearly $\phi(k)$ does not take full advantage of the fact that the decision variables for the first k clinics are fixed. However, $\phi(k)$ can be calculated prior to enumeration and appears to be more efficient computationally than a more complicated bounding function.

A comprehensive experimentation has not been performed because runs on the model have been dictated by the needs of clinic administration and the medical staff in the evaluation of various potential decisions involving the organization of the clinics, facility allocation, and teaching and research schedules. Computational results on requested runs by clinic management, however, have been encouraging. On the average, the clinics had about two alternative schedules that they would accept. Consequently, the integer programs run had 54 to 70 zero-one integer variables. Two runs with 70 variables were solved optimally in 60 and 360 seconds on an IBM 370-165. Twenty other runs with 54 to 57 variables were solved optimally in an average of 7 seconds (12.7 seconds maximum).

It should be pointed out that there are alternative approaches to solving this quadratic programming problem. It is possible to convert the quadratic program to a linear program; for example, see [5, 6, 11]. That formulation was considered early in our project. For the problem at hand, although the LP formulation was possible, we did not see any apparent reduction in solution effort resulting from the Glover-Woolsey formulation. This formulation significantly expands the required number of variables. Further, we had on hand an easily adapted linear 0-1 code which could be tailored to the structure of the quadratic program. This modification was straightforward; the resulting code was only 80 FORTRAN statements and solved the problem with reasonable efficiency. We recognize that this approach would probably not have been the method of choice of an analyst not intimately familar with our code. However, we had deadlines to meet and chose a method which took a minimum amount of development effort. In fact, as the project progressed, we discovered that our approach became increasingly more efficient since the addition of constraints cuts down the solution time dramatically.

Implementation

The implementation of the model was a qualified success. Although relatively straightforward to describe, the final model formulation was derived over a period of more than two years. During that time, in addition to individual meetings with physicians and administrators, several meetings of the entire medical staff were held to discuss the clinic scheduling problem and the model's role in the solution to that problem. These meetings were attended by over 50 physicians and administrators who represented a wide diversity of views. Given this diversity, the fact that a consensus was arrived upon on the validity of the formulation and reasonableness of the data base is particularly encouraging.

At the present time, an optimal schedule is not routinely implemented. However, the model is being used as a planning tool, with potential changes in the clinic schedule being evaluated with respect to the objective function and constraints of the model. Thus, while actual schedules are frequently suboptimal, current planning on the basis of room-hour requirements, cross-consultation flows, and balancing loads on ancillary services is directly attributable to the construction of the model.

References

[1] Balas, E., "An Additive Algorithm for Solving Linear Programs with Zero-One Variables," *Operations Research*, **13** 4 (July-August 1974).

[2] Balinski, M. L., "Integer Programming: Methods, Uses, Computation," *Management Science*, **12** 3 (November 1965).

[3] Ferreira, A. C., and Hodgson, T. J., "An N-Product, Multi-Machine, Lot-size Scheduling Model," *AIIE Transactions*, **5** 3 (September 1973).

[4] Geoffrion, A. M., and Marsten, R. E., "Integer Programming: A Framework and State of the Art Survey," *Management Science*, **18** 9 (May 1972).

[5] Glover, F., and Woolsey, G., "Further Reduction of 0-1 Polynomial Programming Problem to a 0-1 Linear Program," *Operations Research*, **21** 1 (January-February 1973).

[6] Glover, F., and Woolsey, G., "Converting the 0-1 Polynomial Programming Problem to a 0-1 Linear Program," *Operations Research*, **22** 1 (January-February 1974).

[7] Hanan, Maurice, and Kurtzberg, Jerome M., "A Review of the Placement and Quadratic Assignment Problems," *SIAM Review*, **14** 2 (April 1972).

[8] Kunzi, H. P., and Oettli, W., "Integer Quadratic Programming," *Recent Advances in Mathematical Programming*, Edited by Wolfe and Graves, McGraw-Hill Book Co., New York (1963).

[9] Pierce, John F., "Application of Combinatorial Programming to a Class of All-Zero-One Integer Programming Problems, *Management Science*, **15** 3 (November 1968).

[10] Rubin, A. A., and Hammer, P. L., *Quadratic Programming with 0-1 Variables*, Mimeograph Series No. 53, Technion, Israel Institute of Technology (October 1969).

[11] Watters, L. J., "Reduction of Integer Polynomial Programming Problems to 0-1 Linear Programming Problems," *Operations Research*, **15** 6 (November-December 1967).

Dr. Thom J. Hodgson is an Associate Professor of Industrial and Systems Engineering at the University of Florida. He holds BSE (Science), MBA (Quantitative Methods), and PhD (Industrial Engineering) degrees from The University of Michigan. His research interests are in production scheduling and control, Markov renewal programming, and health applications of operations research. Dr. Hodgson is a Senior Member of AIIE and belongs also to ORSA and TIMS.

Dr. Kerry E. Kilpatrick is Director of the Health Systems Research Division and Associate Professor of Industrial and Systems Engineering and Community Health and Family Medicine at the University of Florida. His research interest is in the analysis and design of Health Care Delivery Systems. He received BSE (ME), BSE (Math), MS (Industrial Administration), and PhD Degrees from The University of Michigan, and an MBA from Harvard. As a Senior Member of AIIE Dr. Kilpatrick serves on the Editorial Board of *AIIE Transactions*. He is Chairman-Elect of the Health Applications Section of ORSA and also is on the Editorial Board of the *Journal of Medical Systems*.

Ira M. Longini, Jr., is a PhD candidate in the Biometry and Health Information Systems program at the University of Minnesota, Minneapolis. He is currently working on a dissertation in epidemic modeling with an emphasis on determining optimal vaccination strategies for influenza control. Mr. Longini has a BSIE and MS in operations research from the University of Florida. He is a member of AIIE and Alpha Pi Mu.

Some Network Flow Models and Heuristics for Multiproduct Production and Inventory Planning

JAMES R. EVANS
MEMBER, AIIE
College of Business Administration
University of Cincinnati
Cincinnati, Ohio 45221

Abstract: The focus of this paper is on modelling dynamic, deterministic, multiproduct, multistage production planning problems as single commodity network flow problems. The advantages of such formulations are discussed relative to computation of good integer solutions and continuous solutions with additional constraints. Special results for a single period problem are developed.

■ In this paper we consider dynamic, deterministic, multiproduct, and multistage production-inventory planning problems from a network flow viewpoint. Single product models are easily solved as network flow problems as noted by Zangwill [9], provided that the objective function is linear. If fixed costs are present, a dynamic programming algorithm may be applied to either the single stage or multistage single product problem. The recent book by Johnson and Montgomery [7] provides a thorough discussion of these procedures. The multiproduct model, when viewed as a multicommodity network flow problem with capacitated arcs, is much more difficult to solve since the extreme points of the corresponding linear program are generally non-integer. Thus, efficient single commodity network procedures cannot be used. Bradley [1] discusses these aspects in more detail. The author [3], however, has shown that a special case, the single stage problem, can be formulated as an equivalent single commodity network flow problem which has an integer optimal solution.

In this paper we shall investigate network models for obtaining approximate integer solutions to the problem. The justification for this lies in the fact that the problem is discrete by its very nature, but more importantly, good integer heuristics are computationally more efficient than linear programming approaches for very large problems, especially if the data itself is not exact.

Received January 1976; revised May 1976; August 1976.

Mathematical Formulation

We will use the following notation:

X_{jt}^k = amount of product k produced at stage j in period t

I_{jt}^k = inventory of product k at stage j carried over from period t to $t+1$

P_{jt} = production capacity at stage j in period t

U_{jt} = upper limit on inventory holding from period t to $t+1$

D_t^k = demand for product k in period t

c_{jt}^k = unit production cost at stage j for product k in period t

h_{jt}^k = unit holding cost from period t to $t+1$ at stage j for product k

r = number of products

T = number of periods

N = number of stages .

Mathematically, the problem is

$$\text{P1:}\quad \min \sum_k \sum_j \sum_t [c^k_{jt} X^k_{jt} + h^k_{jt} I^k_{jt}]$$

subject to

$$X^k_{jt} - X^k_{j+1,t} + I^k_{j,t-1} - I^k_{jt} = 0 \qquad \text{for all } t,k;j=1,2,\ldots,N-1 \quad (1)$$

$$X^k_{Nt} + I^k_{N,t-1} - I^k_{Nt} = D^k_t \qquad \text{for all } t,k \quad (2)$$

$$\sum_k X^k_{jt} + \alpha_{jt} = P_{jt} \qquad \text{for all } j,t \quad (3)$$

$$\sum_k I^k_{jt} + \beta_{jt} = U_{jt} \qquad \text{for all } j,t \quad (4)$$

$$X^k_{jt}, I^k_{jt}, \alpha_{jt}, \beta_{jt} \geq 0 \qquad \text{for all } j,t,k$$

Equation (1) is the usual inventory balance equation for each product, stage, and time period. Variables for initial and final inventories are included for the sake of completeness; in many cases these are assumed zero. In (3) and (4) α_{jt} and β_{jt} are slack variables associated with the capacity constraints. We have not established upper or lower bounds for individual X^k_{jt} or I^k_{jt}. However, it will be clear that such constraints may be added with little difficulty. The problem may be viewed as a multicommodity flow problem with capacitated arcs as shown in Fig. 1.

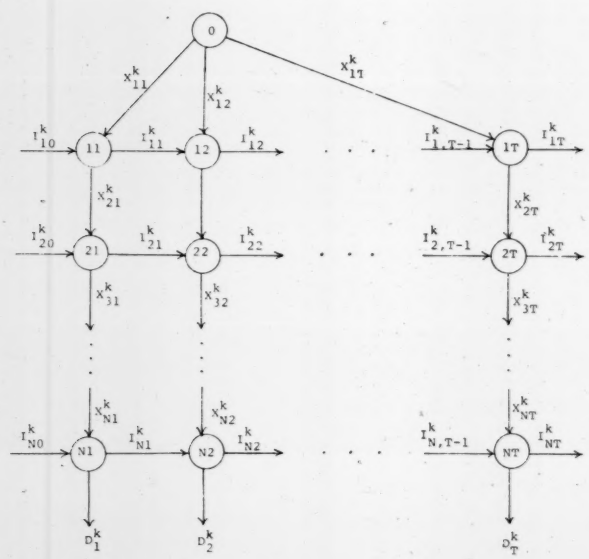


Fig. 1. Multicommodity network analogy.

Any number of multicommodity flow algorithms can be applied to this problem, for instance [5]. However, since the model is only approximate and very large, heuristic methods, particularly those producing integer solutions, often suffice. This is the focus of our discussion.

Single Period Model

In this section we will consider a single period of P1, (Fig. 2). Mathematically, the formulation is

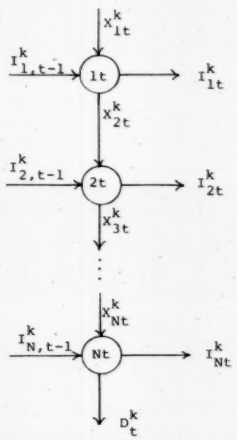


Fig. 2. Single period problem.

$$\text{P2:}\quad \min \sum_{k=1}^{r} \sum_{j=1}^{n} [c^k_{jt} X^k_{jt} + h^k_{jt} I^k_{jt}]$$

subject to

$$X^k_{jt} - X^k_{j+1,t} - I^k_{jt} = -I^k_{j,t-1}. \qquad \text{for all } k;\ j=1,2,\ldots,N-1 \quad (5)$$

$$X^k_{Nt} - I^k_{Nt} = D^k_t - I^k_{N,t-1} \qquad \text{for all } k \quad (6)$$

$$\sum_k X^k_{jt} + \alpha_{jt} = P_{jt} \qquad j=1,2,\ldots,N \quad (7)$$

$$\sum_k I^k_{jt} + \beta_{jt} = U_{jt} \qquad j=1,2,\ldots,N \quad (8)$$

$$X^k_{jt}, I^k_{jt}, \alpha_{jt}, \beta_{jt} \geq 0 . \qquad \text{for all } j,k .$$

In this model, $I^k_{j,t-1}$ are presumed to be known values of inventory entering period t.

In [2] it was shown that a single stage of P1 can be formulated as a single commodity network flow problem. We establish a similar result for P2.

Consider the following set of equations.

$$\sum_k X_{1t}^k + \alpha_{1t} = P_{1t} \tag{9}$$

$$-X_{jt}^k + X_{j+1,t}^k + I_{jt}^k = I_{j,t-1}^k \quad \text{for all } k; \; j=1,2,\ldots,N-1 \tag{10}$$

$$-X_{Nt}^k + I_{Nt}^k = -D_t^k + I_{N,t-1}^k \quad \text{for all } k \tag{11}$$

$$-\sum_k \{ \}_{jt}^k - \beta_{jt} = -U_{jt} \quad j=1,2,\ldots,N \tag{12}$$

$$-\alpha_{jt} + \beta_{jt} + \alpha_{j+1,t} = P_{j+1,t} - P_{jt} + U_{jt} - \sum_k I_{j,t-1}^k \quad j=1,2,\ldots,N-1 \tag{13}$$

$$-\alpha_{Nt} + \beta_{Nt} = -P_{Nt} + U_{Nt} + \sum_k D_N^k - \sum_k I_{N,t-1}^k \tag{14}$$

Equations (9) - (14) define a node-arc incidence matrix of a network. The general network is depicted in Fig. 3. The

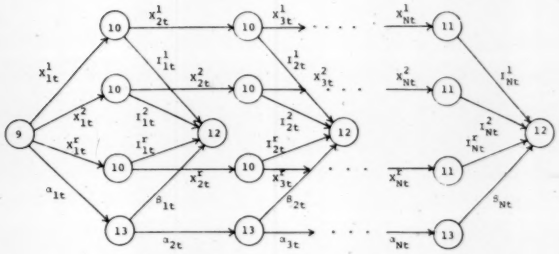


Fig. 3. Single period equivalent network.

node numbers refer to the particular flow conservation constraint (9) - (14) that they represent. The value of the right-hand-side of each conservation equation determines whether that node is a source, sink, or pure transshipment node. A positive value indicates a source, a negative indicates a sink, and a zero value refers to a pure transshipment node.

Theorem 1: The network problem defined by (9) - (14) solves P2.

Corollary The optimal solution to P2 is integer-valued.

These results are analogous to the single stage model formulated as an equivalent network problem in [3]. Theorem 1 is easily proven by performing elementary row operations on Eqs. (5) - (8) to obtain (9) - (14). Linear programming equivalence is maintained and since P2 is a minimal cost network problem, if $I_{j,t-1}^k$, U_{jt}, and P_{jt} are integers, the corollary follows.

Although the network described above solves P2, we shall see that there is a trivial solution. Define, for $j=1,2,\ldots,N$

$$\gamma_{jt}^0 = P_{jt}$$

$$\gamma_{jt}^k = \gamma_{jt}^{k-1} - \max\left\{ D_t^k - \sum_{i=j}^N I_{i,t-1}^k, 0 \right\}$$

$$k=1,2,\ldots,r.$$

We shall assume that $U_{jt} \geq U_{j,t-1}$ for all j. In actuality, this is not a particularly restrictive assumption, since inventory capacity rarely changes from period to period. In many cases U_{jt} will equal $U_{j,t-1}$. This assumption simplifies the following.

Theorem 2: P2 has a feasible solution if and only if, for each $j=1,2,\ldots N$, $\gamma_{jt}^k \geq 0$ for all k.

Proof. Suppose there exists a feasible solution such that for some $1 \leq p \leq r$ and some $1 \leq j \leq N$, $\gamma_{jt}^p < 0$. By definition,

$$\gamma_{jt}^p = P_{jt} - \sum_{k=1}^p \max\left\{ D_t^k - \sum_{i=j}^N I_{i,t-1}^k, 0 \right\}. \tag{15}$$

We assume that

$$P_{jt} < \sum_{k=1}^p \max\left\{ D_t^k - \sum_{i=j}^N I_{i,t-1}^k, 0 \right\}. \tag{16}$$

But the right-hand-side of (16) is simply the demand which cannot be fulfilled by inventory at stages $j, j+1, \ldots, N$ for the first p products. If this exceeds the production capacity at stage j, then clearly there cannot be a feasible solution, contradicting the hypothesis.

To prove the converse, assume that $\gamma_{jt}^k \geq 0$ for all j, k Define

$$X_{jt}^k = \max\left\{ 0, D_t^k - \sum_{i=j}^N I_{i,t-1}^k \right\}.$$

We will show that this yields a feasible solution. Note that

$$\gamma_{jt}^r = P_{jt} - \sum_{k=1}^r \max\left\{ D_t^k - \sum_{i=j}^N I_{i,t-1}^k, 0 \right\}$$

$$= P_{jt} - \sum_{k=1}^r X_{jt}^k \geq 0. \tag{17}$$

Production capacity constraints are clearly satisfied. Next, noting that

$$I^k_{jt} = X^k_{jt} - X^k_{j+1,\,t} + I^k_{j,t-1} \tag{18}$$

we have

$$X^k_{j+1,t} - X^k_{jt} = \max\left\{0, D^k_t - \sum_{i=j+1}^{N} I^k_{i,t-1}\right\} - \max\left\{0, D^k_t - \sum_{i=j}^{N} I^k_{i,t-1}\right\}.$$

Therefore,

$$0 \leqslant X^k_{j+1,t} - X^k_{jt} \leqslant I^k_{j,\,t-1}. \tag{19}$$

From (18) and (19), it follows that $I^k_{jt} \geqslant 0$.

Also, since

$$\sum_k I^k_{jt} = \sum_k [X^k_{jt} - X^k_{j+1,t}] + \sum_k I^k_{j,t-1}$$

and $$\sum_k I^k_{j,t-1} \leqslant U_{j,\,t-1},$$

then $$\sum_k I^k_{jt} \leqslant U_{j,t-1} \leqslant U_{jt}.$$

Therefore $\{X^k_{jt}\}$ is a feasible solution.

Theorem 3: If P2 is feasible, then

$$X^k_{jt} = \max\{0, D^k_t - \sum_{i=j}^{N} I^k_{i,t-1}\}$$

is optimal.

Proof. In the proof of Theorem 2, $\{X^k_{jt}\}$ was shown to be feasible. If $X^k_{jt} < \max\{0, D^k_t - \Sigma^N_{i=j} I^k_{i,t-1}\}$, then demand cannot be met. If $X^k_{jt} > \max\{0, D^k_t - \Sigma^N_{i=j} I^k_{i,t-1}\}$, then these additional units must incur a non-zero holding cost for the next period. Q.E.D.

Intuitively, Theorem 3 states that production should be minimal so as to meet demand exactly. Then $\gamma^r_{jt} = \alpha_{jt}$, the slack variable associated with the production capacity constraints. Note that the Wagner-Whitin property, $X_t I_{t-1} = 0$ need not hold for capacitated problems. Because of the upper bounding constraints, a linear programming basis is not necessarily an arborescence flow. This is true of general multicommodity problems [5].

Multiperiod Models

Although the single period model has a trivial solution, multiperiod models do not, since it may be necessary to produce additional units in order to satisfy demand in later periods. In the remainder of this paper we consider models and procedures for obtaining approximate integer solutions.

An obvious approach is to consider P1 as r independent single commodity network flow problems, initially ignoring the capacity constraints (3) and (4). If each of the single commodity problems are solved and no arc capacities (3) or (4) are exceeded, then clearly the optimum has been attained. If some capacities are exceeded, it is a rather difficult problem getting back to a feasible solution. One could, on the other hand, force feasibility by judiciously allocating P_{jt} and U_{jt} among the r commodities.

In this section we wish to propose a "stronger" relaxation in the sense that a lesser number of constraints of type (3) and (4) are relaxed in the model. The implications of such an approach will be discussed.

Consider Fig. 3. The single period model was constructed using the fact that $I^k_{j,t-1}$ were known inputs to period t. In period $t+1$, I^k_{jt} can be considered as inputs. From this concept, and relaxation of the constraint

$$\sum_k I^k_{jt} + \beta_{jt} = U_{jt}$$

one may adjoin a single commodity network for period $t+1$ to the network for period t. In general, the network is represented by the following set of constraints.

$$\sum_k X^k_{1t} + \alpha_{1t} = P_{1t} \qquad \text{for all } t \tag{20}$$

$$-X^k_{j1} + X^k_{j+1,1} + I^k_{j1} = I^k_{j0} \qquad \text{for all } k;\ j=1,2,\ldots,N-1 \tag{21}$$

$$-X^k_{jt} - I^k_{j,t-1} + X^k_{j+1,t} + I^k_{jt} = 0 \qquad \text{for all } k;\ j=1,2,\ldots N-1;\ t=2,3,\ldots,T \tag{22}$$

$$-X^k_{Nt} - I^k_{N,t-1} + I^k_{Nt} = -D^k_t \qquad \text{for all } k;\ t=2,3,\ldots,T \tag{23}$$

$$-X^k_{N1} + I^k_{N1} = -D^k_1 + I^k_{N0} \qquad \text{for all } k \tag{24}$$

$$-\sum_k I^k_{jT} - \beta_{jT} = -U_{jT} \qquad \text{for all } j \tag{25}$$

$$-\alpha_{j1} + \beta_{j1} + \alpha_{j+1,1} = P_{j+1,1} - P_{j1} + U_{j1} - \sum_k I^k_{j0} \qquad j=1,2,\ldots,N-1 \tag{26}$$

$$-\alpha_{N1} + \beta_{N1} = -P_{N1} + U_{N1} + \sum_k D^k_1 - \sum_k I^k_{N0} \tag{27}$$

$$-\alpha_{NT} - \beta_{N,T-1} + \beta_{NT} = \sum_k D^k_T - P_{NT} - U_{N,T-1} + U_{NT} \tag{28}$$

$$-\alpha_{jt} - \beta_{j,t-1} + \beta_{j,t} + \alpha_{j+1,t} = P_{j+1,t} - P_{jt} + U_{j,t} - U_{j,t-1} \qquad j=1,2,\ldots,N-1;\ t=2,3,\ldots,T. \tag{29}$$

Each equation corresponds to a node, and supplies and demands are determined by the right-hand side. Sources are specified by positive right-hand sides. An example for N=3, T=2, and r=2 is given in Fig. 4. As before, node numbers

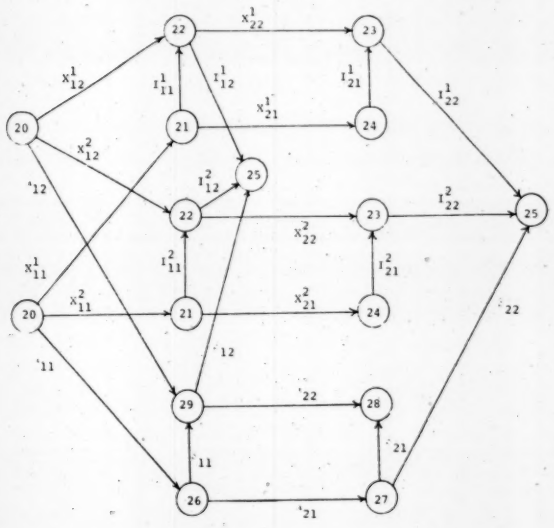


Fig. 4. Example network for N=2, T=2, r=2.

correspond to the constraints (20) - (29). Investigation of this problem will reveal that all constraints of P1 are satisfied except

$$\sum_k X_{jt}^k \leq P_{jt} \qquad j=2,3,\ldots,N \text{ and all } t \tag{30}$$

$$\sum_k I_{jt}^k \leq U_{jt} \qquad t=1,2,\ldots,T\text{-}1 \text{ and all } j. \tag{31}$$

This relaxation is stronger than the total relaxation of (3) and (4) since (3) is satisfied for j=1 and all t; and (4) is satisfied for t=T and all j. Although only N+T constraints out of a total of $2NT$ constraints are enforced, this formulation provides other advantages.

First, solution of this problem yields an objective function value greater than or equal to that of the total relaxation method since more constraints are enforced, and hence, a solution that is closer to feasibility. Secondly, the products are linked together via the network. The significance in this fact lies in the following discussion. Chen and Saigal [2] have developed an algorithm for solving continuous network flow problems with additional linear constraints. One could apply their algorithm in two ways. First, relax (3) and (4) and consider these as "additional constraints." In essence, this is the basis for the generalized upper bounding algorithm of Hartman and Lasdon [5] in solving the general multicommodity flow problem. Or, one could consider the proposed relaxation with (30) and (31) as "additional constraints." Since the proposed relaxation method contains a fewer number of "additional constraints" the working basis would be smaller in general, and one would expect such a procedure to be more efficient.

Heuristic Methods

In [4], Evans and Jarvis developed an out-of-kilter based integer heuristic algorithm for the multicommodity transportation problem using a similar relaxation strategy as described in the previous section. Limited computational experience indicated that very good solutions are obtained. In this section we wish to describe how this technique can be applied to the relaxation of (20) - (29); and also present alternatives for problems with a small number of periods.

The relaxation can easily be solved using the out-of-kilter algorithm [8]. If the resulting solution is feasible then it is clearly optimal. Otherwise, let Γ be the set of arcs whose total flow exceeds capacity. An initial feasible solution can be obtained by allocating P_{jt} or U_{jt} among the various commodities on arcs in Γ. This semi-capacitated problem is then resolved. To each arc (i,j) is associated an updated cost $\bar{C}_{ij}$ that arises from linear programming duality in the out-of-kilter algorithm. A complete treatment of the theory can be found in [8]. $\bar{C}_{ij}$ represents the marginal change in total cost per unit increase in the capacity of arc (i,j). Hence a potential improvement in the current solution exists only if $\bar{C}_{ij} < 0$ and the capacity of that arc can be increased, i.e., it is not equal to P_{jt} or U_{jt}. We emphasize "potential" since a capacity change may force the capacity of some arc not in Γ to be violated, or result in a basis change in which case $\bar{C}_{ij} < 0$ was valid only over a zero range. This may result in a higher cost. Hence, we define a *local optimum* as a solution such that no capacity exchange preserves feasibility and results in a lower cost. The iterative step consists of choosing the arc with the most negative $\bar{C}_{ij}$, determining if a strict improvement exists and if so, the maximum increase in the capacity of that arc. The maximum increase in the arc capacity is determined by one of the following blocking conditions:

(i) No capacity constraint may be violated
(ii) All flows must remain non-negative
(iii) The allocations of P_{jt} or U_{jt} among commodities for arcs in Γ must remain constant.

This may be determined with the out-of-kilter algorithm by finding the breakthrough path that forms a cycle with the arc under consideration. Details are given in [4]. To illustrate the general strategy, consider the problem in Fig. 4, a 2-stage, 2-period, 2-commodity problem. All production costs are 5 and holding costs are 1. Capacities, demands and initial inventories are given in Table 1. In Table 2 we have the optimal linear programming solution (LP) to (1) - (4), and the out-of-kilter solution (OOK) to the relaxed problem (20) - (29). In the relaxed problem, production capacity on arcs (2,1) and (2,2), and inventory capacity on arcs (1,1) and (2,1) are relaxed. In the resulting solution, $X_{21}^1 + X_{21}^2 > P_{21}$; therefore the solution to the relaxed problem is not feasible.

Table 1:

i	j	P_{ij}	U_{ij}	D_i^j	I_{io}^j
1	1	60	15	20	5
1	2	90	15	50	0
2	1	70	10	40	10
2	2	105	15	60	5

Table 2:

i	j	k	$X_{ij}^k(LP)$	$X^k{}_{ij}(OOK)$	$I_{ij}^k(LP)$	$I_{ij}^k(OOK)$
1	1	1	25	0	10	0
1	1	2	35	45	0	0
1	2	1	30	40	0	0
1	2	2	60	50	0	0
2	1	1	20	20	0	0
2	1	2	45	55	0	10
2	2	1	40	40	0	0
2	2	2	60	50	0	0

An initial feasible solution can be obtained by allocating $P_{21} = 70$ among commodities 1 and 2; e.g., setting the capacity on arc $X_{21}^1 = 20$ and on arc $X_{21}^2 = 50$. The problem is then resolved yielding the feasible solution in Table 3. In

Table 3:

i	j	k	X_{ij}^k	I_{ij}^k
1	1	1	15	0
1	1	2	45	5
1	2	1	40	0
1	2	2	50	5
2	1	1	20	0
2	1	2	50	0
2	2	1	40	0
2	2	2	55	0

this out-of-kilter solution, the only negative updated arc cost corresponds to X_{21}^1 and is -1. This indicates a potential improvement. The next step is to increase the capacity on this arc by one unit. This makes the arc out-of-kilter and a flow augmenting path is sought. Resolving, we find the same result, indicating that the previous solution was an alternate optimal solution (in this case, the updated arc cost was valid only over a zero range). Had a strict improvement been found, the capacity on arc X_{21}^1 would be increased as much as possible while preserving (i) - (iii). A negative updated arc cost would then be sought, and the process repeated. Computational results in [4] for multicommodity transportation problems indicate that this method yields solutions that are generally within seven percent of the optimum cost. Since any multicommodity flow problem can be reformulated as a multicommodity transportation problem, such a procedure appears effective.

If good estimates of demand are known for only a few periods in advance, an arc-path formulation provides some advantages. For the example under consideration, application of Theorems 2 and 3 provides the optimal solution for period 1 as follows: $X_{11}^1 = 15, X_{11}^2 = 35, X_{21}^1 = 20, X_{21}^2 = 45$. Given this solution, the optimal policy for period 2 is $X_{12}^1 = 40, X_{12}^2 = 60, X_{22}^1 = 40, X_{22}^2 = 60$. However, the constraint for P_{12} is violated. A rerouting of flow is necessary. A simple solution can be obtained via an arc path formulation.

Start with the optimal flows for period 1 and zero flow for period 2. Let $P_1, P_2, \ldots, P_e$ be an enumeration of all paths from node 0 to node N2 in Fig. 1. Let A be such that

$a_{ij} = 1$ if arc i is on path P_j and = 0 otherwise.

Let y_j^k be the flow of commodity k on path P_j and d_j^k be the sum of the costs of arcs on path P_j. An optimal, feasible solution to the 2-period problem can be obtained by solving

$$\min \sum_k d^k y^k$$

$$[A \quad A \ldots A] \begin{bmatrix} y_2^1 \\ y \\ \cdot \\ \cdot \\ \cdot \\ y^r \end{bmatrix} \leqslant b$$

$$\sum_j y_j^k = D_2^k, \text{ all } k$$

$$y_j^k \geqslant 0$$

where each component of b equals the capacity of the corresponding arc minus the total flow on that arc. In the example we have

d_j^k: path	11	11	10	11	11	10	
Arc	y_1^1	y_2^1	y_3^1	y_1^2	y_2^2	y_3^2	
X_{11}	1	1	0	1	1	0	$\leqslant$ 60-15-35
X_{21}	0	1	0	0	1	0	$\leqslant$ 70-20-45
X_{12}	0	0	1	0	0	1	$\leqslant 90$
X_{22}	1	0	1	1	0	1	$\leqslant 105$
I_{11}	1	0	0	1	0	0	$\leqslant 15$
I_{21}	0	1	0	0	1	0	$\leqslant 10$
	1	1	1				$= 40$
				1	1	1	$= 60$

The optimal solution is $y_1^1 = 10, y_3^1 = 30, y_3^2 = 60$. For this problem, the linear program has 16 constraints and 16 variables; for the arc-path formulation, 8 constraints and 6 variables. For a small number of periods, savings are evident; however, the number of paths grows exponentially with the

number of periods, precluding the use of this approach for more than 2 or 3 period problems.

As a final remark, we wish to point out that single period planning problems with constraints in addition to (5) - (8), for instance resource availabilities, can be effectively handled using the single commodity equivalent model (9) - (14) in conjunction with the algorithm of Chen and Saigal [2]. The computational effort is far less than straightforward application of linear programming.

Acknowledgment

The author wishes to thank the referees for their constructive comments.

References

[1] Bradley, G., "Survey of Deterministic Networks," *AIIE Trans.*, 7, 222-234 (1975).

[2] Chen, S. and Saigal, R., "A Primal Algorithm for Solving A Capacitated Network Flow Problem With Additional Linear Constraints," Bell Laboratories, Homdel, New Jersey (1975).

[3] Evans, J. R., "A Single Commodity Network Model for A Class of Multicommodity Dynamic Planning Problems, Working Paper, Department of Quantitative Analysis, University of Cincinnati.

[4] Evans, J. R., and Jarvis, J. J., "An Improved Formulation and A Heuristic Algorithm for the Integer Multicommodity Transportation Problem," Industrial and Systems Engineering Report Series J-75-14, Georgia Institute of Technology, 1975.

[5] Hartman, J. K. and Lasdon, L. S., "A Generalized Upper Bounding Algorithm for Multicommodity Network Flow Problems," *Networks.* 1, 333-354 (1972).

[6] Johnson, E. L., "Networks and Basic Solutions," *Opns. Res.* 14, 619-623 (1969).

[7] Johnson, L. A., and Montgomery, D. C., *Operations Research in Production Planning, Scheduling and Inventory Control,* Wiley, 1974.

[8] Phillips, D. T. and Jensen, P. A., "Network Flow Optimization With the Out-of-Kilter Algorithm, Part I–Theory," Research Memorandum 71-2, School of Industrial Engineering, Purdue University.

[9] Zangwill, W. I., "A Backlogging Model and a Multi-Echelon Model of a Dynamic Economic Lot Size Production System–A Network Approach," *Man. Sci.* 15, 506-527 (1969).

Dr. James R. Evans is an Assistant Professor of Quantitative Analysis in the College of Business Administration at the University of Cincinnati. He received BSIE and MSIE degrees from Purdue and a PhD from Georgia Tech. His current research interests include theory and applications of network flows, graph theory, and combinatorial optimization. In 1972, Professor Evans received the AIIE-Armstrong Student Award for Excellence. He is a member of AIIE, ORSA, TIMS, MAA, Alpha Pi Mu, Tau Beta Pi, Omicron Delta Kappa, Phi Kappa Phi, and Pi Mu Epsilon.

The Development of an Automated Program for the Optimal Design of Bulk Cushioning Systems

DON McDANIEL
SENIOR MEMBER, AIIE
Advanced Systems Concept Office
US Army Missile Command
Redstone Arsenal, Alabama 35809

Abstract: This paper presents an automated approach to the design of bulk cushioning systems. The developed code can be utilized in package and shipping container design, and other impact absorption applications. A model of impact response is developed and its validity demonstrated through statistical testing. The model is then used as the objective program for a sequential search procedure that searches out the best cushion design.

■ Packaging is a 25 billion dollar a year business in the United States [1]. Almost every item that is grown or manufactured is packaged for shipment to a consumer, whether the consumer is an individual purchasing from the manufacturer or the farm, a retail store stocking inventory, or the military buying from a contractor. The field includes the packaging of cosmetics, pharmaceutical supplies, food products, as well as automobiles, airplane parts, electronic equipment, and guided missiles.

One of the most exciting new materials to enter the packaging arena from the chemical laboratories in the last fifteen years is foamed plastic. This family of versatile, light-weight, low-cost, easily fabricated materials is particularly adapted for use as the cushioning media in packages. The use of foamed plastics in packaging is estimated to be in excess of 100 million dollars a year in the United States and, with the advancing technology, the projected use is expected to double each year for the next few years [2].

The increased use of foamed plastics as cushioning in packages has created the need for a procedure that determines the best cushion design for packages. This research effort addressed itself to that problem. In addition, an automated version of the procedure was developed that is available to package designers for package design.

Received December 1975; revised August 1976.

Modeling the Impact Response of Foam Cushions

The current techniques for cushion design do not produce optimal designs since all the parameters that have a significant effect on cushion performance variations are not included. One particular omission is the temperature variations a package encounters in transportation and storage, even though it has been demonstrated that temperature variations have a significant effect on cushion performance [3]. In developing a model of impact response that is valid, it is necessary to include the parameters that have a significant effect on impact response. A functional relationship of the variables expressed mathematically follows:

$$G = f(\sigma_s, T, \theta, h) \tag{1}$$

where

G = acceleration G-level

σ_s = static stress in psi

T = thickness of cushion in inches

θ = cushion temperature in °F

h = drop height in inches.

The theory of viscoelasticity was utilized to provide a theoretical basis for relating the various parameters in the model. Viscoelastic theory recognizes cushioning materials as possessing mechanical properties that are common to perfect solids and perfect liquids. Various theories have been developed over the past century for describing the behaviour of perfect solids and perfect liquids. Among these, the oldest theories are the classical theory of elasticity and the theory of hydrodynamics. Mathematical models have been formulated for these types of viscoelastic materials, and they are expressed compositely in constitutive equations which, in viscoelastic theory, describe the response of materials to mechanical excitation.

To include temperature effects in the constitutive equation that models impact response, one must consider that portion of viscoelastic theory which accounts for the behavior of materials at varying temperatures. A partial basis is found in a paper on the viscoelastic properties of thermorheologically simple cushioning materials by Cost [4]. The general expression developed by Cost is:

$$\ddot{x}_{max} \approx C_0 \left(\frac{h}{\sigma_s T}\right)^{1/2} + \sum_{i=1}^{N} \frac{C_i h^{1/2}}{(\sigma_s T)^{i+1/2}} \left[\sum_{j=1}^{M} K_{ij}\theta^j\right] \quad (2)$$

where $\ddot{x}_{max}$ is the peak deceleration in G's experienced by a mass impacting a cushion of thickness (T), with a static stress σ_s, from a height of h inches and at a temperature θ, and $C_i (i = 0, 1, \ldots, N)$ and $K_{ij} (j = 1, 2, \ldots, M)$ are constants to be determined by curve fitting procedures.

The viscoelastic model developed by Cost was selected as the most viable candidate in providing a functional relationship of all the variables required in Eq. (1).

Development of the General Model

A research effort was initiated at the Army Missile Command (MICOM) in 1973 to generate data on the cushioning properties of bulk cushioning materials at various temperatures. The initial experimentation was conducted on "Minicel," a cross-linked polyethylene foam material manufactured by Hercules, Inc. [6]. This data base was utilized in validating the Impact Response Models.

The initial model, Eq. (2), proved deficient in representing the nonlinear characteristics of cushion response. An extensive literature search showed it to be the only known model that provides a direct relationship of the required variables and, consequently, a modular technique suggested by Shannon [5] was used to modify the model and construct a valid model of impact response. The relationship of each independent variable (σ_s, T, θ, h) and its effect upon the dependent variable (G-level) was studied and the finalized relationship for each independent variable was then incorporated as shown in Eq. (3)

$$G = C_0 + \sum_{\ell=0}^{S} h^{\ell/2} \sum_{k=0}^{R} \frac{1}{T^{(1/2+k)}} \sum_{j=1}^{N} \theta^j \sum_{i=0}^{M} C_{ijk\ell} (\ell n\, \sigma_s)^i. \quad (3)$$

This General Model incorporates each of the variables in the manner prescribed in the modular modeling effort and incorporates the following characteristics:

1) The dynamic cushioning curves of G-level versus static stress generated by the model are U shaped. This is consistent with cushioning theory.
2) The predicted G-levels increase with increased drop height. The higher drop heights incorporate more energy into the system which would increase the energy levels experienced by the falling body.
3) The predicted G-levels increase with reduced cushion thickness. The G's experienced by a falling body are dependent on shock pulse duration, and a thinner cushion would allow less shock pulse time and an accompanying increase in G-levels.
4) Temperature effects induce lateral shifts in the dynamic cushioning curves. This is consistent with the phase shift function concept of viscoelastic theory. It is also intuitively consistent in that reduced temperatures would be expected to stiffen the cushioning material and require an increased stress level for comparable shock attenuation.
5) The dynamic cushioning curves generated as output from the General Model form a series of curves that are nested within progressive values of thickness and drop height for all possible temperature conditions and static stress conditions.

A sensitivity analysis was performed using the Minicel data. S is set at 1, R is set at 1, N is set at 3, M is set at 2, and the Minicel Model takes the following form:

$$G = C_0 + \sum_{\ell=0}^{1} h^{\ell/2} \sum_{k=0}^{1} \frac{1}{T^{(1/2+k)}} \sum_{j=1}^{3} \theta^j \sum_{i=0}^{2} C_{ijk\ell} (\ell n\, \sigma_s)^i. \quad (4)$$

The resulting Minicel Model can be written as follows:

$$G = -8.39 + \frac{3.54\,\theta\,(\ell n\, S)^2}{T^{1/2}} - \frac{15.31\,\theta\, h^{1/2}\, \ell n\, S}{T^{3/2}}$$

$$+ \frac{3.34\,\theta\, h^{1/2}\,(\ell n\, S)^2}{T^{3/2}} + \frac{207.49\,\theta\, h^{1/2}}{T^{1/2}} - \frac{50.35\,\theta h^{1/2} \ell n\, S}{T^{1/2}}$$

$$+ \frac{1.43\,\theta\, h^{1/2}\,(\ell n\, S)^2}{T^{1/2}} - \frac{6.70\,\theta^2\, \ell n\, S}{T^{1/2}} - \frac{54.66\,\theta^2\, h^{1/2}}{T^{1/2}}$$

$$+ \frac{11.62\,\theta^2\, h^{1/2}\, \ell n\, S}{T^{1/2}} - \frac{1.30\,\theta^3}{T^{1/2}} + \frac{2.08\,\theta^3\, \ell n\, S}{T^{1/2}}$$

$$- \frac{0.23\,\theta^3\,(\ell n\, S)^2}{T^{1/2}} - \frac{0.40\,\theta^3\, h^{1/2}}{T^{3/2}} + \frac{0.61\,\theta^3\, h^{1/2} \ell n\, S}{T^{3/2}}$$

$$- \frac{0.09\,\theta^3\, h^{1/2}\,(\ell n\, S)^2}{T^{3/2}} - \frac{0.87\,\theta^3\, h^{1/2}\, \ell n\, S}{T^{1/2}} - \frac{233.77\,\theta}{T^{3/2}}$$

$$+\frac{28.30\,\theta\,\ell n\,S}{T^{3/2}}-\frac{49.68\,\theta^2}{T^{3/2}}+\frac{26.32\,\theta^2\,\ell n\,S}{T^{3/2}}-\frac{6.07\,\theta^2(\ell nS)^2}{T^{3/2}}$$

$$-\frac{6.15\,\theta^3\,\ell n\,S}{T^{3/2}}+\frac{1.07\,\theta^3\,(\ell n\,S)^2}{T^{3/2}}+\frac{3.94\,\theta^3\,h^{1/2}}{T^{1/2}} \qquad (5)$$

where $\theta = \frac{^{\circ}F + 460}{100}$ and S = static stress in psi × 100.

This model can be used to predict impact response for Minicel cushioning systems. The model specificity is that it is 95% reliable when used within the ranges of the independent variables which are as follows:

1) h = drop height from 12 through 30 inches.
2) σ_s = static stress range from 0.03 to 5.0 psi.
3) θ = temperature from -65° to 160°F.
4) T = thickness of cushion from 1 through 3 inches.

The form of the General Model as given in Eq. (5) was used in the validation of the model for the cross-linked polyethylene foam Minicel material. If additional precision had been required the values of S, R, M and N could have been increased, which may be necessary with other materials, but gave only a minimal improvement in precision here. It was apparent, however, that increases in M, which incorporates σ_s with exponentials over 2, is in general not worthwhile. Also, increases in N above 3 that incorporate θ are of marginal value in improved precision. The changes in R and S that affect the exponentials of thickness and drop height should be explored initially if additional precision is required in the modeling of a particular material.

To fully verify the model as a predictor, a data base independent of that used to generate the model must be utilized. Three such data bases were available: a data base of of 1-, 2-, and 3-inch thickness Minicel material at -65°,70° and 160°F and at a 27-inch drop height; a data base of 4- and 5-inch Minicel material at the usual drop heights.

To determine statistically how well the model fits a set of independent data, Box and Draper [9] suggest it is appropriate to investigate the bias and variance of the predictor. Two statistical tests can be utilized for this purpose. One test, a test of means, determines whether there is a bias in the predicted values of the model when compared to actual data values. The other test, a test of variances, determines if the variations of the predicted values inherent in the model are comparable with the variations in experimental values.

The test of means and variances were conducted on the Minicel Model, using the 27-inch drop height data and the 4- and 5-inch data samples. The results are given in Table 1 and show the following:

1) The 27-inch drop height data are within the range of the independent variables used in the Minicel Model and the null hypothesis cannot be rejected in the test of means or test of variances. This indicates that there is no significant difference between the means and variances of the data and the values predicted by the model. The model appears to be a statistically valid predictor of these data.

2) The test results for the 4- and 5-inch data show that the Minicel Model gives good predictions of the data means at the 4-inch thickness. The test of means of the 4-inch showed no significant difference but the test on the 5-inch showed a significant difference, which is not surprising, because the model is projecting 2 inches beyond its data base range. The test of variances shows that the model variance at the 4-inch thickness is significantly larger, but this is not true at the 5-inch level. The model dispersion is comparable to the dispersion of the 5-inch data.

Table 1: Tests of Minicel Model as a predictor.

	Minicel material (α= 0.05)		
	27-inch Drop Height	4-inch Thickness	5-inch Thickness
Test of Means			
Number of cells in the data base (N)	99	156	156
Standard deviation of the cells (S)	20.14	8.63	8.56
t_m	1.84	-1.45	-2.61
t_{table}	±1.96	±1.96	±1.96
Decision on H_o: $E(\Delta_j) = 0$	cannot reject	cannot reject	reject
Test of Variances			
Variance of the data, σ_d^2	111.63	26.41	105.35
Variance of the model, σ_m^2 (Residual Mean Square)	114.39	114.39	114.39
F_{calc}	1.02	4.33	1.08
F_{table}	$F_{1511,198}$=1.17	$F_{1511,312}$=1.15	$F_{1511,312}$=1.15
Decision on H_o: $\sigma_d^2 = \sigma_m^2$	cannot reject	reject	cannot reject

The results of the tests of means and variances show that the model was not completely acceptable at either the 4-inch or 5-inch thickness. However, the tests indicate that model means were comparable with the data at the 4-inch level and the variances were comparable at the 5-inch level.

Optimization

Once a valid math model of impact response has been developed for a particular material such as Minicel, it is then possible to formulate an automated optimization procedure that will provide the best design for that particular cushioning material. The developed procedure can best be described

as a constrained sequential direct search technique in that the mathematical model of a material such as Minicel is used directly as the objective function [10].

The best cushioning system is one that provides the necessary shock isolation to the packaged item at a minimum cost. Since the cost of a cushioning system is dependent upon the amount of cushioning material, it follows that the best cushioning system will employ the minimum thickness of cushion needed. The optimal point on a dynamic cushioning function is the minimum G-level which provides the maximum protection for that particular thickness of cushioning and therefore, it is the identification of this optimal point that must be determined in selecting the best cushion system.

A different dynamic cushioning function is required to depict cushion performance for each drop height, thickness, and temperature of the cushion; and the optimal point is different for each function. When one considers the many types of cushioning materials, it can be seen that a very large library of functions is required to present even the most likely conditions of drop height, temperature, and thickness of a few candidate cushioning materials.

The impact response is computed directly from the values of the independent variables thereby precluding the need for a library of dynamic cushioning curves. Also, the development of valid mathematical models of impact response provides a vehicle for the application of optimization techniques. This model can be reconstituted as an objective function and explored for optimality. This can be done by defining cushioning system design requirements in terms of the external environment and the amount of exposure the packaged item can withstand.

The specification of the external environment must contain a definition of those aspects of the environment that have an impact upon cushion system design. This would include a quantitative measure of the magnitude of the maximum shock pulse to which the package will be exposed, which is usually specified in terms of drop height (h). Also, since temperature has a significant effect on impact response, the specification of the external environment should include the range of temperatures to which the package will be exposed (θ_{min}, θ_{max}).

The specification of the survivability of the protected item is given in terms of its ability to withstand shock. The maximum shock pulse, in G's, that the item can withstand is given as the fragility level of the item (GL_{max}).

These considerations concerning the G-levels and temperature are incorporated into the model as input values that are utilized to constrain the optimal search to feasible solutions. GL is identified as the fragility level of the protected item and the temperature parameter, θ, must consider the range of temperatures of the external environment θ_{min} to θ_{max}. This can be accomplished utilizing the superimposed dynamic cushioning function format with the extremes of the temperature range superimposed upon the ambient.

The drop height input (h) can be determined based on published testing standards for packages. Other parameters, such as the fragility level of the item to be protected (GL_{max}) and the temperature range (θ_{min}, θ_{max}), are usually specified in the packaging requirements. For example, if the cushioning system is to be designed for a military container for a missile, the fragility level of the protected item (GL_{max}) will be specified as part of the hardware specifications in the missile system design criteria. The temperature range (θ_{min}, θ_{max}) is defined in the missile system requirements and the drop height is specified in various Military Standards depending on total container weight. Equation (1) can now be written as follows:

$$GL_{max} = f[\sigma_s, T, (\theta_{min}, \theta_{max}), h] \tag{6}$$

where GL_{max}, (θ_{min}, θ_{max}), and h are inputs to the equation. The optimization technique searches out the best design from this expression, and the solution is expressed in terms of σ_s and T.

The procedure searches upon the functional relationship expressed in Equation (7) and with the inputs introduced, reduces to the selection of the minimum thickness of cushion that will perform satisfactorily at the determined static stress condition.

The search is initiated on a minimum thickness cushion (1 inch) at the drop height (h) and minimum temperature (θ_{min}) stipulated in the design requirements. These values are input into the objective function and the search is conducted across the static stress spectrum at the temperature extremes, θ_{min} and then θ_{max}, to determine the feasibility of meeting the stimulated GL_{max}. If the search of the acceptable static stress range on the minimum cushion thickness is successful, the answer is output as a feasible solution. If it is not successful, the thickness is incremented ½ inch and the search repeated.

The direct search program output is given in the form of superimposed dynamic cushioning functions such as the typical one given in Fig. 1. In Fig. 1, the design objective

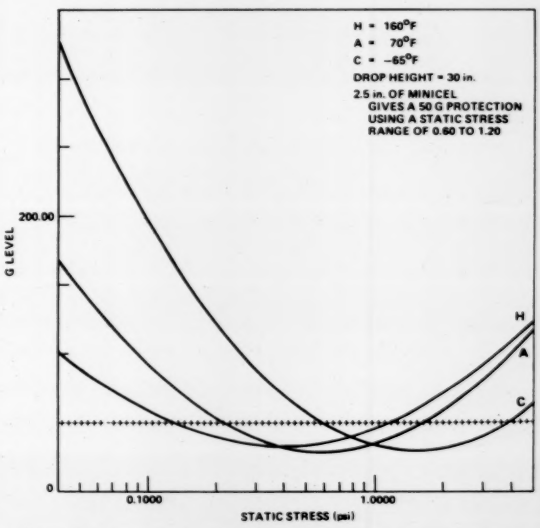


Fig. 1. Search procedure output, 2½ inch Minicel.

function has the temperature range of -65° to 160°F and a 30-inch drop height. The fragility level was 50 G's which is indicated with a horizontal line. The minimum thickness is determined to be 2.5 inches and the feasible static stress range is found to be 0.60 to 1.20 psi. One of the inherent advantages of using math models is seen here in the capability of determining G-level response at thicknesses other than those in the data base. In this instance a 2.5-inch thickness of cushion was best; tests were not conducted at this thickness in the data base. This advantage can also be seen in Fig. 2 which is another output of the optimization program where the temperature range has the non-data base values of -20° and 120°F. The minimum thickness is 2.0 inches and the stress range is 0.55 to 0.80 psi to give 50-G protection. Comparison of Fig. 1 with Fig. 2 demonstrates the effect of temperature on optimal cushion design. When

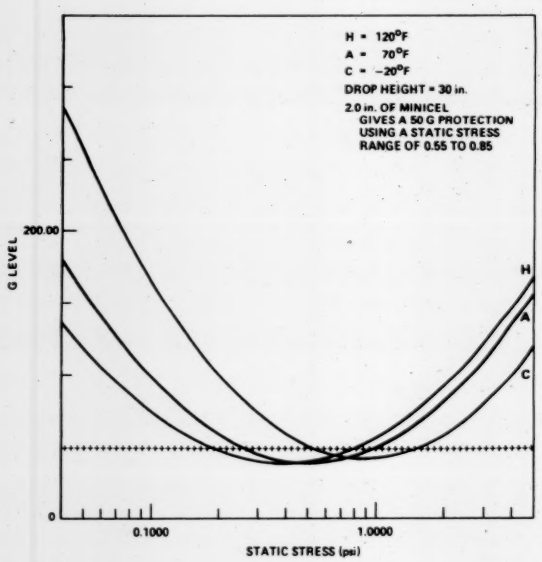


Fig. 2. Search procedure output, 2-inch Minicel.

the temperature range was relaxed from the extremes of -65° through 160°F to -20°F through 120°F and all other exogenous variables kept the same, the thickness of cushion required for 50-G protection was reduced from 2.5 inches to 2 inches.

Conclusions

There has been limited use made of bulk cushioning in shock isolation systems for commercial and military packaging. One reservation that designers have had regarding bulk cushioning is an inability to reliably predict cushioning performance.

One result of this research is the development of a model of bulk cushioning response that provides a valid predictor of bulk cushioning materials impact response. The developed model is predicated on viscoelastic theory and incorporates the effect of temperature, stress, drop height, and thickness of cushion upon the impact response of a cushioning system. This General Model provides the basic underlying structure of impact response of any one of the many bulk cushioning materials used for shock isolation. The Minicel Model is one such model that was constructed using the General Model as the basic underlying structure. The Minicel Model showed high correlation with the actual data within the ranges of the variables and also shows promise as a predictor beyond the variable ranges. The functional form of the model provides the advantage of determining impact response at non-tested levels of the variables and precludes the need for a library of dynamic cushioning functions for all the different combinations of conditions.

The development of a valid model of impact response of bulk cushioning materials defines the relationships and interrelationships of the variables in response to impact. Using this basis, it was possible to employ a search technique to determine the best cushion design. This program accepts the design requirements for a shock isolation system and computes the best design for each cushioning material in the data base, if one exists.

Models are being formulated on four additional cushioning materials. These are linear polyethylene foam (Ethafoam) in two densities, 2 lb/ft^3 and 4 lb/ft^3; 4 lb/ft^3 polyester urethane; and 3 lb/ft^3 polyether urethane. These models, when validated, will be incorporated into the Container Optimization Library. Their validity will be, of course, subject to the lot to lot variations that must be reasonably controlled during the manufacturing process, if the behavior of the tested samples is to be representative of the material in general. It is also important to realize that the cushion design output from the optimization program is only one of several determinations that must be made in the preliminary design of a complete shipping container, or other end item; and that prototype testing of the complete design is highly desirable in determining the acceptability of the design.

The cost effectiveness of this research effort can be assessed in terms of the benefits that will accrue from the use of the optimal cushion design program by package designers throughout industry and the government. Better designs and cost avoidance in superfluous fabrication and testing that effect even a small percentage improvement represent sizable savings in a $100 million market. Secondly, the improved predictability this program brings to bulk cushion design will stimulate the use of bulk cushions in substitution for the more expensive and complex mechanical suspension systems utilized in shipping containers.

Finally, and perhaps most consequential from a cost avoidance standpoint, there will be a reduction in the failure of package cushioning systems. The cost of the packaged items is usually several orders of magnitude greater than the packaging materials. A small percentage reduction in damage to the billions of dollars of packaged items takes on astronomical proportions.

References

[1] Simms, William C. (Ed), *Modern Packaging: Encyclopedia and Packaging Planning Guide*, **47**, 12, McGraw-Hill, Inc., p. 12 (December 1974).

[2] "How Big is Plastics in Packaging," *Modern Plastics*, McGraw-Hill Publishing Co., N.Y., pp. 98-99 (May 1967).

[3] McDaniel, Don, *A Study of Impact Test Effects Upon Foamed Plastic Containers*, US Army Missile Command, Redstone Arsenal, Alabama, Report No. RF-TR-71-2 (1971).

[4] Cost, Thomas L., *Dynamic Response of Container and Container Cushion Structures*, Athena Engineering Company, Northport, Alabama, Technical Report No. 74-003 (October 1974).

[5] Shannon, R. E., *Systems Simulation: The Art and Science*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1975).

[6] Wyskida, Richard M. and Wilhelm, Mickey R., *Temperature Sensitive Dynamic Cushioning Function Development and Validation for Hercules Minicel Thermoplastic Foam*, The University of Alabama in Huntsville, Huntsville, Alabama, UAH Research Report No. 159 (MICOM Report No. RL-CR-75-1) (September 1974).

[7] Duncan, Acheson J., *Quality Control and Industrial Statistics*, Richard D. Irwin, Inc., Homewood, Illinois (1965).

[8] Daniel, Bailey E., *Probability and Statistics – Models for Research*, John Wiley & Sons, Inc. New York (1971).

[9] Box, G. E. P. and Draper, N. R., "A Basis for the Selection of a Response Surface Design," *Journal of American Statistical Association*, **54** (1959).

[10] McDaniel, Don, *Modeling the Impact Response of Bulk Cushioning Materials*, US Army Missile Command, Redstone Arsenal, Alabama, Technical Report RD-75-16 (1975).

Dr. Don McDaniel is a Research Scientist with the US Army Missile Command Research, Development, and Engineering Laboratory located at Redstone Arsenal, Alabama. He received a Bachelor's degree in Engineering from Ohio State University and a MS and PhD in Industrial Engineering from the University of Alabama in Huntsville. Dr. McDaniel has been engaged in research and development of containers for Army Missile Systems since 1966. He is a Senior Member of AIIE, and a Registered Professional Engineer in Ohio and Alabama.

Price Adjusted Single Sampling with Quadratic Indifference

JOSEPH W. FOSTER
SENIOR MEMBER, AIIE
Department of Industrial Engineering
Texas A&M University
College Station, TX 77843

OSCAR T. PERRY
U.S. Army Armament Research and Development Command
Dover, New Jersey 07801

Abstract: The purpose of this paper is to expand the application of Price Adjusted Single Sampling, PASS, to include quadratic indifference. The concept of PASS is first reviewed. Secondly, the statistical background of PASS with quadratic indifference is formulated. Finally a step by step procedure for implementing a PASS plan is presented along with an example.

■ Traditional acceptance sampling was developed out of the need to overcome inherent shortcomings of the 100% inspection technique. Its oirginal purpose remains essentially unchanged: to provide an economical method of inspecting material and a mathematically sound procedure for arriving at a decision to accept or reject material.

One of the inherent processes of traditional sampling techniques is lot rejection. Its purpose is to influence the producer in his choice of p, the incoming lot fraction defective. However, the rejection process constitutes a principal disadvantage to these techniques because it results in many additional costs, e.g., reshipment and reinspection costs. Regardless of arrangements made in contracts, experience reveals that these additional costs will ultimately be absorbed by both the consumer and producer.

The purpose of Price Adjusted Sampling is to eliminate the lot rejection process on the premise that it is possible to do so yet still retain the power to influence the producer who controls p. The concept of Price Adjusted Sampling was first introduced by Foster [5] and was further developed by Foster and Perry [6] into a specific set of tabled plans belonging to Price Adjusted Single Sampling, PASS, with linear indifference. This paper uses the same general concepts developed for sampling with linear indifference and applies them to quadratic indifference. The result is a procedure for developing plans which, together with the plans requiring linear indifference, provide a wider range of flexibility in the selection of a suitable price adjusted plan.

Received December 1974; revised June 1975, April 1976, August 1976.

Review of Linear Indifference

The underlying concept used in all price adjusted sampling plans is that of "consumer indifference." Originating in the field of economics, it is used to model the behavior of the "rational" consumer, [11], toward a number of combinations of the quantities of two or more commodities. For example, suppose a consumer is concerned only about two products, x and y, and that he can in every instance decide whether he prefers one combination of x and y quantities to another, or whether either combination is equally acceptable to him; i.e., he is "indifferent between" the two combinations.

The consumer's attitude may be graphically illustrated by a curve similar to that shown in Fig. 1. The points A, B, C represent combinations of products x and y which the consumer has decided are equally acceptable to him. The actual curve, then, extending beyond these points, represents his indifference curve of the two commodities.

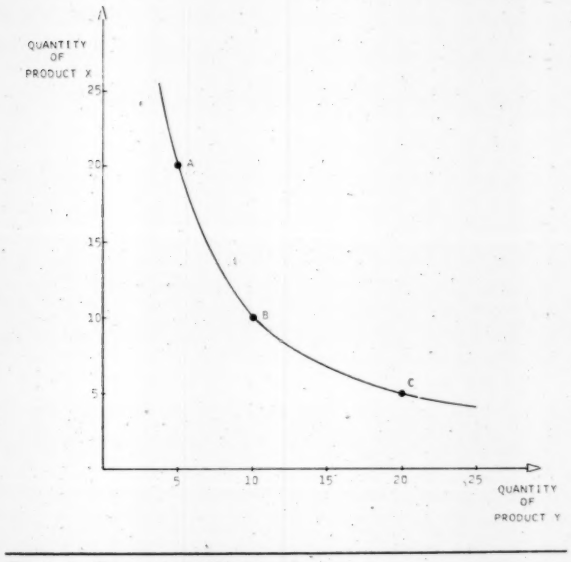


Fig. 1. An Economic Indifference Curve

Figure 2 represents a typical linear indifference curve for PASS, wherein each combination of lot price and quality level of the lot is equally acceptable to the consumer. The equation of the curve is defined as the indifference function $h(p)$. The constants N and P are defined, respectively, as the lot size and the average price per acceptable unit the consumer is willing to pay. Thus, in the linear case, $h(p)$ may be written as

$$h(p) = NP(1 - p) \tag{1}$$

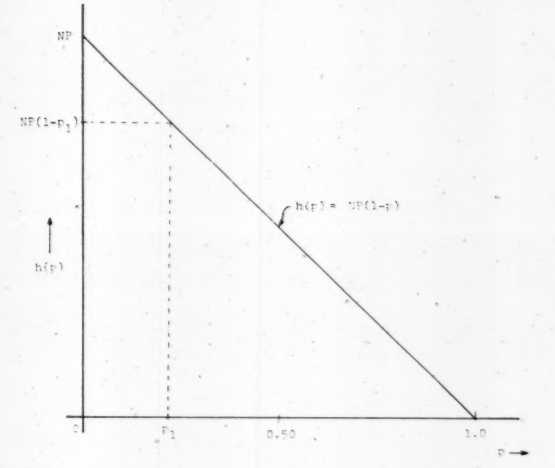


Fig. 2. A PASS Linear Indifference Curve.

Using Fig. 2, it can easily be shown that the consumer is in fact "indifferent to the quality level." At any given quality level, say p_1, the total number of good units contained in the lot is expressed as $N(1 - p_1)$. From Fig. 2, the consumer pays $NP(1 - p_1)$ for the lot. The price paid per acceptable unit, then, designated $C(p)$, is obtained from the ratio of these two quantities, i.e.,

$$C(p_1) = \frac{h(p_1)}{N(1 - p_1)} = \frac{NP(1 - p_1)}{N(1 - p_1)} = P,$$

which is exactly the price he is willing to pay for an acceptable unit. Hence, it follows that regardless of the value of p for a given lot, the consumer will always pay the amount P for an acceptable unit, and thus he is not concerned about the quality of any particular lot.

In order to define an indifference curve for a particular situation, the cost consequences of various quality levels must be evaluated. One method would be to estimate the total manufacturing cost to produce the units from one lot of received material at zero fraction defective. Then make similar estimates at other levels of fraction defective. The indifference curve could then be defined such that the total production cost plus the procurement cost be equal for all levels of fraction defectives. Therefore, as a producer the total cost would be the same for any level fraction defective. Hence, you would be indifferent to the quality level.

The discussion thus far has assumed that the quality level of a lot is known. Since this is generally not the case, it is necessary to resort to sampling techniques resulting in Price Adjusted Single Sampling, or PASS. The statistical objectives of PASS, regardless of the shape of the indifference curve, are to first, base the payment of a lot on the inspection results of a sample taken from the lot; and second, insure that the expected value of his payment will be that specified by his indifference curve. This then makes the consumer indifferent to the quality level in the long run.

The Need for Indifference Curves other Than Linear

While the linear indifference plans described above yield a constant price P per acceptable unit for all p, circumstances may exist in which the consumer is not satisfied with this concept. For example, if lots of purchased assembly parts are excessively defective, it may be necessary for the consumer to shut down his assembly operations. Consequently, he may want the amount he pays for a lot to reflect his increased cost.

A quadratic indifference function, as will be seen in this paper, will provide this further flexibility in PASS plans by allowing the price paid per acceptable unit, $C(p)$, to vary in different ways over the range of p. For example, $C(p)$ can be made to remain nearly a constant for smaller values of p, say between 0 and 0.3 and then decrease slowly until p approaches 1.0, wherein it decreases at a considerable rate. This particular variation in $C(p)$ may satisfy the consumer's dilemma noted in the previous example.

Statistical Model for PASS

The statistical theory of PASS is based on a sample size of n drawn randomly from a population of N units. The number

of defects found in the sample is denoted by x. The process of selecting x defects out of n draws from a population of N units which are p fraction defective follows the hypergeometric distribution. However, the model for PASS will assume that the ratio of n to N is sufficiently small such that the binomial approximation to the hypergeometric is valid. It should be noted that numerical methods may be employed to solve specific problems if the binomial approximation is not valid in specific situations.

The first step in constructing a PASS plan is to derive an expression for the total lot price which first is a function of only the number of defects in the sample; and secondly adjusts the price based upon sample results. If P_T denotes the total lot price, let

$$P_T = N[P - g(x)], \tag{2}$$

where $g(x)$ is a decrementing function in x which effectively reduces the price per lot as x increases. Note that to accomplish this, $g(x)$ must itself increase as x increases.

Recall that while the PASS indifference curve represents the price the consumer is willing to pay for a lot, assuming p is known, Eq. (2) represents the actual price he pays for the lot. The consumer may be provided with long-range indifference by requiring that the expected value of the total lot price be equal to the indifference function, $h(p)$. That is,

$$E[P_T] = E\{[P - g(x)]N\} = h(p). \tag{3}$$

Equation (3) will be used to determine the decrementing function $g(x)$, which assumes a unique form for a given type of indifference curve, and which depends only on x since p is an unknown parameter.

For the quadratic indifference curve such as that illustrated in Fig. 3, $h(p) = NP(A_0 + A_1p + A_2p^2)$. Substituting this into Eq. (3) and expanding, the result is:

$$E\{[P - g(x)]N\} = NP(A_0 + A_1p + A_2p^2)$$

$$P - E\{g(x)\} = P(A_0 + A_1p + A_2p^2)$$

$$E\{g(x)\} = P(1 - A_0 - A_1p - A_2p^2). \tag{4}$$

Consider

$$g(x) = B_0 + B_1x + B_2x^2. \tag{5}$$

The constants B_i; $i = 0, 1, 2$, may be evaluated by noting that if x is distributed binomially, $E[x] = np$ and $E[x^2] = np(1 - p) + (np)^2$, where n is the sample size. By expanding Eq. (4) and equating powers of coefficients of p, the coefficients of the decrementing function are found to be

$$B_0 = P(1 - A_0)$$

$$B_1 = \frac{PA_2 - PA_1(n - 1)}{n(n - 1)} \tag{6}$$

$$B_2 = \frac{PA_2}{n(1 - n)}$$

Substituting Eq. (5) into (2), with values given in Eq. (6), P_T becomes:

$$P_T = NP\{1 - [K_0 + K_1(n)x + K_2(n)x^2]\} \tag{7}$$

where

$$K_0 = 1 - A_0,\ K_1(n) = \frac{A_2 - A_1(n - 1)}{n(n - 1)} \text{ and } K_2(n) = \frac{A_2}{n(1 - n)}.$$

Equation (7) represents the lot price equation for quadratic indifference, and its use will become evident throughout the remainder of this paper. It should be noted once again that Eq. (7) has accomplished both objectives of PASS; namely that, 1) the lot price is adjusted according to sampling results; and, 2) its expected value is equal to $h(p)$, the indifference function.

The remaining quantity necessary to complete the plan is the sample size, n, which, as in the case of linear indifference, may be obtained from either of two expressions. The first of these provides protection to the producer as follows: if the fraction defective is at some predetermined acceptable quality level, p_1, then the probability that the consumer pays a price per nondefective unit less than or equal to a specified lower limit, L, should be as close to without exceeding some specified value, α. That is,

$$Pr\{\text{Price per nondefective unit} \leqslant L \mid p = p_1\} \leqslant \alpha. \tag{8}$$

In this expression, α may be considered the producer's risk since it is the probability that the producer receives a price lower than L at quality level p_1.

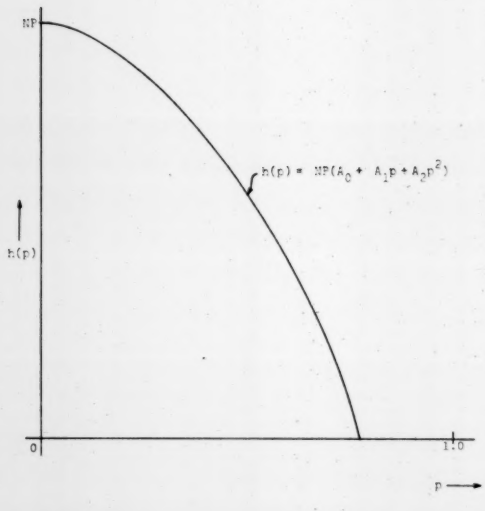


Fig. 3. A PASS Quadratic Indifference Curve.

The second expression provides protection to the consumer as follows: if the fraction defective is at some predetermined quality level, p_2, which may or may not be considered acceptable, then the probability that the consumer pays a price per nondefective unit greater than or equal to a specified upper limit, U, should be as close to without exceeding a specified value, β. That is

$$Pr\{\text{Price per nondefective unit} \geq U | p = p_2\} \geq \beta. \quad (9)$$

In this expression, β may be considered the consumer's risk since it is the probability that the consumer pays a price greater than U at quality level p_2.

While it would seem desirable to protect both the producer and consumer a minimum n will be determined by either Equation 8 or 9 hence both may not specify a protection level. If both are stated the one requiring the largest n could be used thus giving the other party more protection than desired. In any case the range of both over a wide range of p should be analyzed before using PASS.

The Problem

The above considerations lend to the following problem definitions:

Find the minimum value of n which satisfies either Eq. (8) or (9). Equations (8) and (9) may be rewritten as Eqs. (10) and (11), respectively.

$$Pr\{Q(x) - \frac{L}{P}(1-p_1) \leq 0\} \leq \alpha \quad (10)$$

$$Pr\{Q(x) - \frac{U}{P}(1-p_2) \geq 0\} \leq \beta \quad (11)$$

where $Q(x) = Ao + \frac{A_1 x}{n} - \frac{A_2 x}{n(n-1)} - \frac{A_2 x^2}{n(n-1)}$

A_0, A_1, A_2 are constants.

The equations may be solved for n by trial and error. First consider the case where A_2 is negative, i.e., $h(p)$ is concave downward. In this case

$$Pr\{Q(x) - \frac{L}{P}(1-p_1) \leq 0\} = Pr\{X \leq X_1\} + Pr\{X \geq X_2\} \quad (12)$$

where X_1 and X_2 are the roots of

$$Q(x) - \frac{L}{P}(1-p_1) = 0, X_1 \leq X_2$$

and

$$Pr\{Q(x) - \frac{U}{P}(1-p_2) \geq 0\} = Pr\{X_1 \leq X \leq X_2\} \quad (13)$$

where X_1 and X_2 are the roots of

$$Q(x) - \frac{U}{P}(1-p_2) = 0, X_1 \leq X_2.$$

The solution procedure is as follows:

Step 1: Assume $n = 1$.

Step 2: Evaluate $Q(x)$ and final X_1 and X_2.

Step 3: Knowing n, X_1 and X_2, calculate the probability that the random variable X is in the region specified by Eq. (12) or (13) whichever is applicable.

If A_2 is positive $h(p)$ is concave upward. The regions in which the random variable X must fall to satisfy the probability statements are reversed, e.g.,

$$Pr\{Q(x) - \frac{L}{P}(1-p_1) \leq 0\} = Pr\{X_1 \leq X \leq X_2\} \quad (14)$$

and

$$Pr\{Q(x) - \frac{U}{P}(1-p_1) \leq 0\} = Pr\{X \leq X_1\} + Pr\{X \geq X_2\}. \quad (15)$$

The trial and error solution is similar to the above except care must be taken to use the correct region for the random variable X.

Operating Curves

As previously stated, it is usually of interest to show how payments will vary over the range of p rather than at just two points, p_1 and p_2. To satisfy this need, PASS Operations Curves are developed. We first find the minimum n which will satisfy both Eqs. (10) and (11). For this value of n the expected price per non defective unit, $C(p)$, is calculated according to:

$$C(p) = E\left[\frac{P_T}{N(1-p)}\right] = E\left[\frac{P - g(x)}{(1-p)}\right]$$

where $g(x)$ is defined in Eq. (5). Further, an "upper," $U(p)$, and "lower" $L(p)$, price per nondefective unit is (corresponding to α and β, respectively) defined as solutions to the equations

$$L(p) = \frac{P \cdot Q(k_1)}{1-P_1} \quad (16)$$

$$U(p) = \frac{P \cdot Q(k_2)}{1-P_2} \quad (17)$$

where k_1 and k_2 are the positive roots of $Q(x) - (L/P)(1-p_1) = 0$ and $Q(x) - (U/P)(1-p_2) = 0$, respectively.

A necessary condition for Eq. (16) and (17) is that both $Q(x) - (L/P)(1-p_1) = 0$ and $Q(x) - (U/P)(1-p_2) = 0$ have at most one positive root. This is always the case if $h(p)$ is concave downward but may not be the case otherwise. If there are two positive roots then one will be greater than n in order to satisfy the number "3" restriction in Table 1.

In that case substitute X_1, the smallest positive root, for k_1 and k_2 in Eqs. (16) and (17) respectively. A typical Operating Curve is shown in Fig. 4. Note $C(p)$, the expected price per nondefective unit, is provided in the figure as a point of reference for the upper and lower limits of risk.

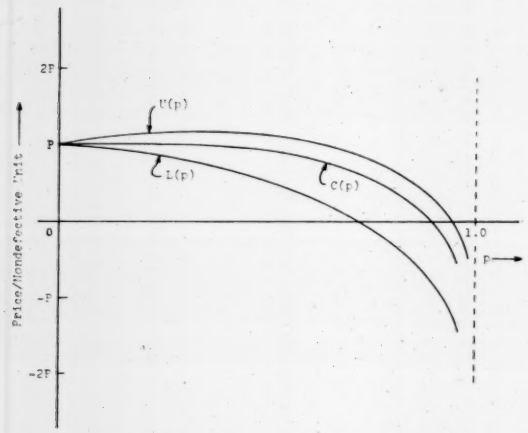


Fig. 4. A typical operating curve for a PASS plan with Quadratic indifference.

It should be noted that for some values of p the expected payments become negative. This should not be disturbing for even in conventional procurement practices a supplier may lose money on a particular lot with an unusually high fraction defective. This is typically the result of remedial action required by the consumer. The producer has then in fact received a negative payment for that lot. PASS allows the negative payment but defines beforehand the extent of penalty to be incurred, thus eliminating the need for negotiations.

A comprehensive coverage of PASS with linear indifference along with indifference curves for comparison is found in [6].

Restrictions on Indifference Curve Parameters

Because of the general nature of the suggested indifference curve, i.e., $A_0 + A_1 p + A_2 p^2$, certain values of the parameters lead to unrealistic consequences and hence should be restricted from use. Unlike the linear indifference case, the restrictions are not necessarily obvious. Table 1 summarizes the restrictions and their associated reasons. For convenience, the table distinguishes between two cases: first, $h(p)$ concave upward, and secondly, concave downward.

The typical sequence of activities required to install a PASS plan starts with a development of an indifference curve by the consumer. This curve should reflect economic issues such as cost of disposing of defective material, etc., if significant. If such a curve violates the restrictions stated above it does not necessarily mean that the curve is invalid only that it should be reviewed in light of the comments associated with the violated restriction. It may be that the curve is valid but not quadratic.

Table 1: Restrictions on A_i, the indifference function coefficients.

Restriction	Reason
CASE I: $h(p)$ Concave Downward, $g(x)$ Concave Upward	
1. $A_2 < 0$	Mandatory for $h(p)$ to be concave downward
2. $A_1 < \frac{A_2}{n-1}$	Necessary if $g(x)$ is to have a positive slope over range $0 \leq x \leq n$. Since n is not known at time $h(p)$ is selected, let $n = 25$ for purpose of making a decision; if $A_1 < A_2/24$, assume that final value of n will also satisfy this restriction.
3. $A_0 = 1$	Not mandatory; is used because consumer most likely will not pay more than NP if p=0; likewise, producer most likely will not accept less than NP if $p = 0$.
4. $4A_2 + 4A_1 \leq -4 \leq A_2 + 2A_1$	Also not mandatory; however, produces a reasonable curve for $h(p)$. The p-intercept, $h(p) = 0$, occurs in range $0.5 \leq p \leq 1.0$.
5. $A_1^2 \geq 4A_2A_0$	Necessary to prevent imaginary roots of $h(p)$. For CASE I. with $A_0 = 1$ and $A_2 < 0$, this restriction is always met.
CASE II: $h(p)$ Concave Upward, $g(x)$ Concave Downward	
1. $A_2 > 0$	Mandatory if $h(p)$ is to be concave upward.
2. $A_1/A_2 < -2$	This restriction has been approximated. Its purpose is to force $g(x)$ to have a positive slope in the range $0 \leq x \leq n$
3. $A_1 < 0$ $0 \leq 4A_2A_0 \leq A_1^2$ $A_0 \geq (L/P)(1-p_1)$ $A_0 \geq (U/P)(1-p_2)$ $A_1 + A_2 \leq -1$	These are all necessary in order to produce positive, non-complex roots of both P_T and $h(p)$ so that solutions may be obtained for Equations 14 and 15. (Only one of the restrictions on A_0 applies, depending on whether the plan is to protect the producer or the consumer.)
4. $A_0 = 1$	Same as CASE I.

NOTE: As can be seen from above, the use of CASE II has been extremely limited by numerous restrictions. Therefore, except in unique circumstances, it will be advantageous to use the CASE I. form of $h(p)$ rather than the CASE II form.

Example

The following example will illustrate the procedure for formulating and implementing a PASS plan. Suppose that a producer and consumer have agreed to sample lots under a PASS plan with quadratic indifference. They have further decided that the plan will protect the producer with $\alpha = 0.05$ for $p_1 = 0.10$. The remaining constants for the plan are P = \$1.00, L = \$0.94, and N = 1000. The Case I form of

indifference function chosen for this plan is $h(p) = NP[1 - 0.5p - 1.5p^2]$. Note that this $h(p)$ satisfies the constraints listed in Table 1.

All that is needed to complete the plan is the sample size, which will be obtained by the trial and error procedure previously described using Equation 10. From $h(p)$ the A's are found:

$$A_0 = 1,\ A_1 = -0.5,\ A_2 = -1.5.$$

Also
$$\frac{L}{P} = 0.94.$$

The solution for this problem is $n = 34$ since

$$Q(x) = 1 - \frac{.5x}{n} + \frac{1.5x}{n(n-1)} - \frac{1.5x^2}{n(n-1)}$$

which for n=34 has roots

$$x_1 = -16.84 \text{ and } x_2 = 6.84$$

hence from Eq. (12)

$$Pr\{Q(x) - \frac{L}{P}(1-p_1) \leqslant 0\} = Pr\{x \leqslant -16.84\} + Pr\{x \geqslant 6.84\}.$$

Since x is a non-negative integer random variable, the last expression is

$$Pr\{x \geqslant 7\} = .048.$$

This is as close to $\alpha = .05$ as is possible without exceeding it.

An operating curve for this PASS plan may be derived by solving Eq. (16) and (17) for given values of α and β; for this example, $\alpha = \beta = 0.05$. Points for the producer's risk curve, $L(p)$, the consumer's risk curve, $U(p)$, as well as for $C(p)$, the expected price per non-defective unit, are all provided in Table 2, and the curve itself is shown in Fig. 5. It is noteworthy that the payment limits are narrow for small p but increases rapidly as p increases beyond 0.1.

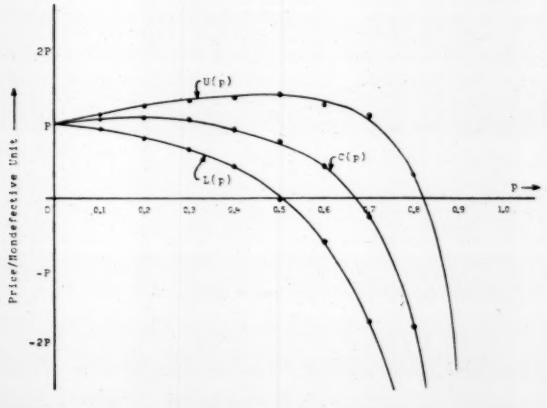


Fig. 5. Operating curve for example plan.

Table 2: Calculated points for PASS operating curve.

P	$L(p)$	$U(p)$	$C(p)$
0	1(P)	1(P)	1(P)
.1	0.934	1.11	1.04
.2	0.808	1.21	1.05
.3	0.632	1.29	1.02
.4	0.436	1.35	0.93
.5	-0.034	1.38	0.75
.6	-0.636	1.24	0.40
.7	-1.718	1.08	-0.28
.8	-4.005	0.28	-1.8

However, since the operating region is usually $p \leqslant 0.04$, this characteristic does not prevent application.

The plan, therefore, is implemented by

1) taking a sample of 34 from a lot of 1000;

2) inspecting to determine x; and

3) paying $P_T = NP[1 - (.0134)x - (.00134)x^2]$ for the lot, where $N = 1000$ and $P = \$1.00$.

Conclusions

To summarize, the basic quantities necessary to complete a PASS plan with quadratic indifference are n, the sample size, and $h(p)$, the indifference function. The first step is to perform a detailed analysis of the restrictions associated with the A_i coefficients in $h(p)$. This is done with the aid of Table 1. Thus, assuming for now that A_i's are known, the next step is to establish who the plan is to protect: if the consumer, β, U, and p_2 must be specified, or if the producer L and p_1 must be specified. Regardless of the choice it is suggested that the Operating Curve be determined before implementing the plan.

Next, values for P and N must be determined which, in general, will reflect the physical properties of the manufacturing process. Finally, $g(x)$ and n are calculated, respectively, from Eq. (5) and from either Eq. (9) or (10).

A PASS plan is implemented as follows:

1) take a sample of size n from a lot of N units;
2) inspect to determine x, the number of defects in the sample; and
3) pay $P_T = NP(1 - [K_0 + K_1(n)x + K_2(n)x^2])$ to the producer for the lot.

The primary purpose of applying PASS is the elimination of rejecting lots. Among the advantages of this elimination are: 1) The loss of time between the rejection of one lot and acceptance of its replacement is eliminated, 2) The cost of reshipping the rejected lot is saved, 3) Since reinspection and reshipment of lots is eliminated, the producer is free of any artificial costs when determining his optimum cost to quality-level trade-off.

Other advantages of PASS not related to the rejection process are:

1) Since the consumer is indifferent to the quality level, any policing action is unnecessary.
2) A quality incentive is automatically given to the producer.

The salient disadvantage of PASS is the difficulty which will be encountered in convincing a supplier and consumer to inspect and pay under such a plan. It would likely require the establishment of an independent bonded company to perform the inspection function. Another disadvantage is the difficulty in identifying an indifference curve. In spite of these difficulties, the authors feel that many situations exist in which PASS may be profitably applied.

References

Periodicals:

[1] Dodge, H. F., "Notes on the Evolution of Acceptance Sampling Plans, Part I," *Journal of Quality Technology*, **1** 2, pp. 77-88 (April 1969).

[2] Dodge, H. F., "Notes on the Evolution of Acceptance Sampling Plans, Part II," *Journal of Quality Technology*, **1** 3, pp. 155-162 (July 1969).

[3] Dodge, H. F., "Notes on the Evolution of Acceptance Sampling Plans, Part III," *Journal of Quality Technology*, **1** 4, pp. 225-232 (October 1969).

[4] Dodge, H. F., "Notes on the Evolution of Acceptance Sampling Plans, Part IV," *Journal of Quality Technology*, **2** 1, pp. 1-8 (January 1970).

[5] Foster, Joseph W., "Price Adjusted Sampling Plan," *International Journal of Production Research*, **8** 4 (1970).

[6] Foster, Joseph W., and Perry, Oscar T., "Price Adjusted Single Sampling with Linear Indifference," *Journal of Quality Technology*, **4** 3, pp. 134-144 (July 1972).

[7] Rocloffs, Robert, "Acceptance Sampling Plans with Price Differentials," *The Journal of Industrial Engineering*, **18** 1 (January 1967).

Books:

[8] Aiken, H. H., and Staff of the Computation Laboratory, *Tables of the Cumulative Binomial Probability Distribution*, Harvard University Press, Cambridge, Mass. (1955).

[9] Bowker, A. H., and Lieberman, G. J., *Engineering Statistics*, Prentice-Hall, Englewood Cliffs, N.J. (1959).

[10] Duncan, A. J., *Quality Control and Industrial Statistics*, Richard D. Irwin, Homewood, Illinois (1965).

[11] Fox, K. A., *Economics Analysis for Public Policy*, Iowa State College Press, Ames, Iowa (1958).

Oscar T. Perry is a General Engineer in the Product Assurance Directorate of the U.S. Army Armament Research and Development Command in Dover, New Jersey, where he is responsible for assessing the reliability and safety of Army nuclear weapon systems. Mr. Perry's research interests are in statistical techniques in quality control, maintainability and production. He has a bachelor's degree in electrical engineering from the University of Arizona and master's degree in industrial engineering from Texas A&M University. He received the Meritorious Award from Picatinny Arsenal and is a member of state and national societies of Professional Engineers.

Dr. Joseph W. Foster, III, Professor of Industrial Engineering at Texas A&M University, is currently conducting research in the areas of quality control, reliability, maintainability and production systems. He holds bachelor's degrees in both mechanical and industrial engineering from Southern Methodist University, master's from Lehigh University, and Doctor of Engineering from the University of Oklahoma. A Senior Member of AIIE, Dr. Foster is the Institute representative on the Board of Directors for the Annual Reliability and Maintainability Symposium. He was elected to membership in Pi Tau Sigma, Sigma Xi, and Alpha Pi Mu honorary fraternities, and is a member of the American Society of Quality Control. He formerly taught at the University of Oklahoma and was a research engineer with Western Electric Company.

Centralized Scheduling and Priority Implementation Heuristics for a Dynamic Job Shop Model

ROSSER T. NELSON
UCLA
Los Angeles, CA 90024

CHARLES A. HOLLOWAY
Stanford University
Stanford, CA 94305

RUBY MEI-LUN WONG
Stanford University
Stanford, CA 94305

Abstract: The problem considered is the scheduling of a job shop with job due dates, intermittent job arrivals, and statistical processing times. Centralized scheduling uses a sequence of static problems for generating priorities at review times. A multi-pass heuristic program, which has proven effective in earlier research, is applied to the up-dated static scheduling problem at each review time. A procedure is proposed for implementing priorities on the shop floor between review times. The procedure is expressly designed to integrate the scheduling of newly arriving jobs to modify the schedule. In simulation experiments using tardiness statistics for evaluation, centralized scheduling and the proposed implementation procedure proved to be an extremely effective combination. Comparison with another procedure that gives the centralized schedule precedence over new arrivals indicates the importance of the implementation procedure when periodic centralized scheduling is used in a dynamic situation.

■ Job shop scheduling has been the focus of a substantial amount of research over the last two decades (for a review of much of this work see [1]). An explanation for this activity is the combination of the theoretical challenge posed by the combinatorial nature of the problem and the wide range of practical applications. Much of the work has concentrated on problems with a static arrival process (all jobs available at the beginning of a scheduling period) and deterministic processing times. The motivation for considering this class of problems is that realistic job shop scheduling problems, characterized by dynamic (intermittent) arrivals and statistical processing times, might, for practical reasons, be modeled as a sequence of static problems.[1] Other research has concentrated on development of scheduling procedures directly applicable to job shops with dynamic arrivals and stochastic processing times. These procedures are characterized by dispatching rules based on local information. The use of only local information allows intermittent arrivals and stochastic variations in processing times to be directly incorporated into the priority calculation.

In [3] the authors investigated the use of schedules generated using static arrivals and deterministic processing times when actual processing times are statistical in nature. The schedules were generated using expected processing times; the simulated implementation of the schedule on the shop floor used actual processing times taken from a specified processing time distribution for each operation. The experiments compared centralized schedule generation with four dispatching rules. From the standpoint of possible application, the key results were that the relative performance of the scheduling procedures on job tardiness criteria was largely independent of the variance of the processing time distributions and that, for a wide variety of conditions, schedules generated centrally using expected processing times outperformed those generated using dispatching rules. The results also indicated the importance of the procedure used for implementing centrally generated priorities on the

[1] Peters [7] has applied this idea to a three machine flow shop, using a branch-and-bound algorithm for minimizing maximum flow time to solve the static problem.

Received May 1975; revised April 1976; revised August 1976.

This research was principally supported by the Office of Naval Research under Contract N00014-69-A-0200-4004, Task NR-047-003, partially by the Western Management Science Institute under a grant from the Ford Foundation, and partially by the Graduate School of Business, Stanford University. Reproduction in whole or in part is permitted for any purpose of the United States Government.

The authors wish to acknowledge the helpful comments of a referee on a previous draft of the paper.

shop floor using actual processing times. Two implementation procedures led to quite different results in the static problem.

This paper focuses on the development and testing of a scheduling procedure for *dynamic* problems with *statistical* processing times. Guided by the results of the earlier research, the proposed centralized priority generation procedure uses expected processing times. In view of the earlier results regarding implementation in the static problems, two implementation procedures are evaluated under varying shop conditions. The implementation options represent two identifiable strategies for using the centralized schedule in a dynamic environment. These options are compared with a dispatching rule suited for direct implementation in a dynamic problem. Based on previous research, the minimum SLACK rule[2] was chosen as representative of an effective due-date oriented priority rule utilizing only local information. The purpose of this paper is to investigate the potential usefulness of centrally generated schedules based on static arrivals and deterministic processing times and the sensitivity of the results to implementation schemes.

The Static Arrivals, Deterministic Processing Times Problem

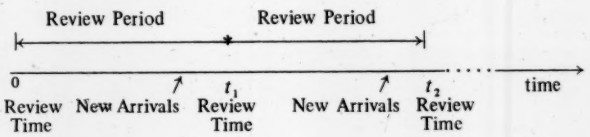
A prerequisite to such an investigation is a procedure for obtaining good solutions to the problem with static arrivals, deterministic processing times and due dates. Although the problem can be formulated as an integer programming problem with a variety of performance measures for the objective function, its size precludes solutions for even relatively small job shop models. Reference [2] describes a procedure, HSP, which has proven to be computationally effective for job shops with static arrivals, deterministic processing times and due dates. (See Appendix A for a brief overview of the procedure.)

The application of HSP requires that provision be made for the case where the specified problem is infeasible (i.e., it is not possible to meet the due dates with the given processing times and job precedence requirements). Previous computational experience has indicated that an efficient strategy is to relax the problem constraints to obtain a problem that is known to be feasible, and then iteratively reapply violated constraints to seek improved solutions. The strategy of relaxation can be applied in a number of ways with the choices taking into consideration the criteria used by management for schedule evaluation. References [5] and [6] describe relaxation strategies for situations in which the resource constraints or job precedence constraints can be relaxed, as through the application of overtime or multiple transports of partial job lots. Reference [4] describes an interactive program that can be used when implicit tradeoffs exist among a number of criteria. For the purpose of the experiments in this paper, we assume that the goal is to minimize maximum tardiness. The relaxation procedure uses the completion times of jobs in the schedule obtained from the SLACK priority rule to establish a set of relaxed due dates for initial use with HSP. These relaxed due dates are then tightened on each iteration to guarantee that maximum tardiness is decreased each time HSP finds a feasible solution. A claim of optimality for the procedure is precluded by the fact that HSP is not guaranteed to find a feasible solution if one exists. This requires that we rely on empirical comparisons with other alternatives in evaluating performance. In addition to the evidence in [2] on the effectiveness of HSP, the iterative procedure used here has previously been applied [5] to eight test problems with an average of 48% improvement in maximum tardiness over the SLACK schedules.

[2]SLACK chooses from jobs in queue that job with the smallest value of [job due date−current time−remaining processing time].

Priority Generation and Implementation for the Dynamic Problem

Use of solutions to static problems in a dynamic problem offers two primary options. The first is to update the system and reschedule each time a new job arrives. In instances where jobs only arrive infrequently, this may be a good strategy and the dynamic aspects of implementation disappear since every new arrival signals a new static problem. In this case schedules for the current static problem are implemented. The second option is to use periodic review times as suggested by the diagram below.



A sequence of review times t_i determine a sequence of review periods which may be fixed or variable in length. At each review time the actual state of the system is updated to reflect current shop status. This would include any impending jobs with known arrival times during the subsequent review period. The current shop status, which includes expected processing times for the operations, is the information assumed to be available for making scheduling decisions for the next review period.

The procedures investigated in this paper use the solution to the static arrivals, deterministic processing times problem to develop a priority list of the jobs known at the review time for each machine. Each priority list includes the sequence of operations scheduled on the machine and the expected arrival time of each operation at the machine. Operations that arrive at the machine during the period that are not on the priority list generated at the beginning of the period are placed on a new job list for the machine. The priority order within this new job list is based on expected SLACK, i.e., highest priority to the operation with the smallest value of [job due date−current time−expected remaining processing time on job].

We consider two procedures for implementing these priorities:

Implementation IP1

When a scheduling decision is to be made on a machine, select from among the operations that are available the operation that is highest on that machine's priority list. If there is no operation available from the priority list, schedule the available operation that is highest in priority on the new job list. IP1 confines attention to operations that are available and selects among these in accordance with (a) the priority list, or (b) the new job list. Thus, IP1 does not permit enforced idle time in the schedule, i.e., a machine is never held idle to wait for a nonavailable operation when there is any work available. This corresponds to the nondelay implementation strategy investigated in [3].

Implementation IP2

Let t = the time at which a scheduling decision is to be made.

Let t_e = the expected arrival time of the highest (unfinished) priority operation on the machine's priority list.

At time t, compute the expected slack S_p of the highest (unfinished) operation on the machine's priority list and the expected slack S_N of the highest (unfinished) operation on the machine's new job list.

Case 1 ($S_p \leqslant S_N$ *or no operations on new job list*)

Schedule the highest operation on the machine's priority list if it is available at time t.

Otherwise:

i. If the shortest expected processing time among available operations is $\leqslant (t_e - t)$, schedule the available operation with the largest expected processing time $\leqslant (t_e - t)$.

ii. If the shortest expected processing time among available operations is $> (t_e - t)$, schedule the operation on the new job list with smallest expected processing time. If the new job list is empty, do not schedule an operation to begin at time t.

Case 2 ($S_N < S_p$ *or no operations on priority list*)

Schedule the highest operation on the machine's new job list at time t.

IP2 differs from IP1 in two respects. First, it is designed to integrate the scheduling of old and new jobs. Second, it allows for delays and selective use of enforced idle time in the schedule. The first step is to use expected slack as a criterion for deciding whether the priority list generated at the beginning of the period or the new job list should receive primary attention in the scheduling decision being made. The former case (Case 1) assigns the highest priority job from the priority list if it is available at time t. If it is not available, the expected time until that operation's arrival ($t_e - t$) is used to attempt to schedule an available operation to reduce the enforced idle time that would be incurred if the machine were left idle until the arrival of the highest priority operation. Step i attempts to make an assignment such that the expected time that the machine will be available is before, but as close as possible to, the expected arrival time of the highest priority operation. If such an assignment is impossible, step ii attempts to assign the operation from the new job list such that the expected time that the machine will be available is as close as possible to the expected arrival time of the highest priority operation. If there is no operation on the new job list, enforced idle time is incurred on the machine until the next operation arrives and a new scheduling decision is made. In Case 2 where the initial step indicates that the new job list should receive primary attention at time t, the highest priority operation from the new job list is scheduled to begin processing at time t. If the scheduling has proceeded to a point where there are no operations remaining on a machine's priority list from the original jobs, operations are scheduled for new jobs on a minimum expected slack basis.

Although there are several details of the two implementation schemes for which alternative choices are readily identified, they are designed to represent the "natural" implementation of two extreme strategies. IP1 corresponds to a non-delay strategy in which centrally scheduled operations always have precedence over new arrivals. IP2 corresponds to a delay strategy in which priority between centrally scheduled operations and new arrivals depends upon the SLACK of the highest priority operation in each category. Results with these strategies as well as suggestions for future research are discussed in upcoming sections.

Illustrative Example

We conclude this section with a detailed example to illustrate the procedure. The data for the example is as follows:

Job	Arrival Time	Routing and Processing Times	Due Date
1	0	D(7,9) A(3,3) B(2,2) E(1,1)	26
2	0	C(20,25) A(2,2) B(2,3)	48
3	0	C(4,3) F(3,4) E(10,10)	34
4	0	A(10,10) F(3,4) B(1,1) D(8,8)	44
5	0	E(1,1) B(3,2) C(6,5) D(1,1) E(1,1)	24
6	0	E(22,25) A(3,2)	50
7	0	E(5,5) D(7,6) A(13,11)	50
8	5	B(9,11) F(1,1) A(6,6) D(1,1)	39
9	15	B(1,1) A(1,1) C(6,4) F(8,8)	47
10	25	F(9,5) B(3,5)	49
11	35	F(3,2) B(11,10) D(10,12)	83
12	45	F(8,10) B(3,4)	67

Letters A through F are machine identifiers. For each operation, the first number in parentheses is the expected processing time. The second number is the actual processing time sampled from a binomial distribution with mean (np) equal to the expected processing time and with the binomial parameter $p = 0.5$. The due date of each job was set equal to arrival time plus two times the total expected processing time of the job. Jobs 1 through 7 are available at time 0 and are assumed to be the only jobs known of at the beginning of the scheduling period.

We summarize the results of applying the priority generation procedure to the example at time 0:

Priority Generation at t = *0*

The SLACK schedule obtained using the due dates, routings, and expected processing times for jobs 1 through 7 results in maximum tardiness of 10. Relaxed due dates were established from the SLACK schedule to initiate the iterative procedure with HSP for seeking improved solutions for maximum tardiness. This led to a schedule with maximum tardiness of 0.

The Gantt chart for the HSP schedule is given in Fig. 1, together with the corresponding machine priority lists. The sequence of job numbers on each machine is simply the sequence from the Gantt chart. The expected arrival time t_e for each operation is the completion time of the previous operation on the job in the Gantt chart.

Machine A		Machine B		Machine C	
Job Number	t_e	Job Number	t_e	Job Number	t_e
4	0	5	1	3	0
1	7	4	13	5	4
2	30	1	13	2	0
7	14	2	32		
6	41				

Machine E		Machine E		Machine F	
Job Number	t_e	Job Number	t_e	Job Number	t_e
1	0	5_1	0	3	4
7	6	7	0	4	10
5	10	3	7		
4	14	5_2	15		
		1	16		
		6	0		

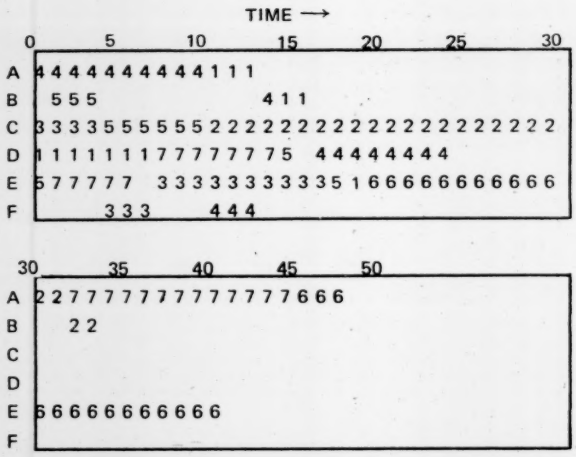


Fig. 1. Gantt Chart for Schedule to be Implemented.

Priority Implementation:

The priorities were implemented with actual processing times using procedures IP1 and IP2. Scheduling with the minimum SLACK single-pass priority procedure[3] was included for comparison. The resulting tardiness statistics are given in Table 1. The IP2 Gantt chart that resulted in zero tardiness is included as Appendix B.

Table 1: Tardiness statistics.

Tardiness Measure	Implementation Procedure		
	SLACK	IP1	IP2
Percent of jobs tardy	33%	33%	0%
Total tardiness	28	44	0
Maximum tardiness	12	18	0

Evaluative Experiments

In this section we summarize the design and results of experiments used for initial evaluation of the procedure described in Section II. Flow charts, listing, and documentation of the (GPSSV) simulation program are given in reference [8].

Simulated Environment. The problem employed for evaluation consisted of 6 machines with 7 jobs available at the beginning of the scheduling period. The data for these jobs is given in Appendix C. One replication of the simulation consisted of one review period and continued until all original jobs and all new jobs that arrived during the review period were completed. New arrivals were generated throughout to prevent a bias in the tardiness statistics due to deteriorating load near the end of the run. The characteristics of new jobs were generated internally by the simulator in accordance with prescribed parameters and distributions (see Appendix C for details).

Experiments. The experimental variables employed were (1) the number of new job arrivals during the review period, (2) the processing time distributions, and (3) the priority implementation procedure.

(1) The ratio N = number of new jobs/number of original jobs was used to explore the sensitivity of the performance of the implementation procedures to the number of new arrivals during the review period. This is viewed as an important factor because it relates to the spacing of the review times in a dynamic problem. The three values of N employed are 0, 3/7, and 1; the

[3]The SLACK priority was computed with expected processing times as would be necessary on the shop floor.

first case being the static problem with only the original jobs. These values span the range from 100% to 50% of the jobs being known by HSP in centralized scheduling.

(2) Actual processing times were generated as the operations were assigned to machines in the simulation. Three distributions were employed:
   (i) a uniform distribution over the range 80% to 120% of expected processing time for the operation.
   (ii) a binomial distribution with $p = 0.5$, symmetric about the mean (np) equal to expected processing time for the operation.
   (iii) a binomial distribution with $p = 0.1$, mean (np) equal to expected processing time, and skewed toward longer processing times.

(3) The priority implementation procedures used were SLACK, IP1 and IP2. SLACK was chosen since it can easily be applied on the shop floor and because it exhibited the best over-all tardiness performance among the single-pass dispatching rules in earlier experiments [3] with static problems and statistical processing times. The HSP program was used with the data for the original jobs to generate the machine priority lists for IP1 and IP2. These priorities were input data to the simulator.

A complete factorial design of the 27 factor combinations was employed. Since the processing times were random variables, 50 independent replications were made for each factor combination.[4]

Results. The tardiness statistics reported were the percentage of jobs tardy and the mean, variance, and maximum tardiness. The complete data, as tabulated in [8], is summarized in Figs. 2 through 5. Figure 3 includes 95% confidence intervals for mean tardiness.

The most notable result is the superior performance of IP2 under all conditions tested and for all four tardiness criteria. This confirms that the implementation strategy is important when using solutions to static problems in a dynamic environment. Previous research on problems with $N = 0$ (see [3]) demonstrated that centrally generated schedules, properly implemented, can be effective over a range of problems and processing time distributions. This research indicates that centrally generated schedules, properly implemented (e.g., IP2), can be effective over a range of $0 \leqslant N \leqslant 1$. This means that centrally generated schedules with review times in which as many as 50% of the jobs arrive during the review period may significantly outperform local dispatching rules. Information similar to that contained in Figs. 2 through 5 could be used by a manager to decide on the proper length of time between reviews by trading-off costs of updating and generating the centralized schedule with tardiness costs.

[4]The simulation program was validated manually using the example in Section II. X^2 tests were used to validate the processing time generator separately.

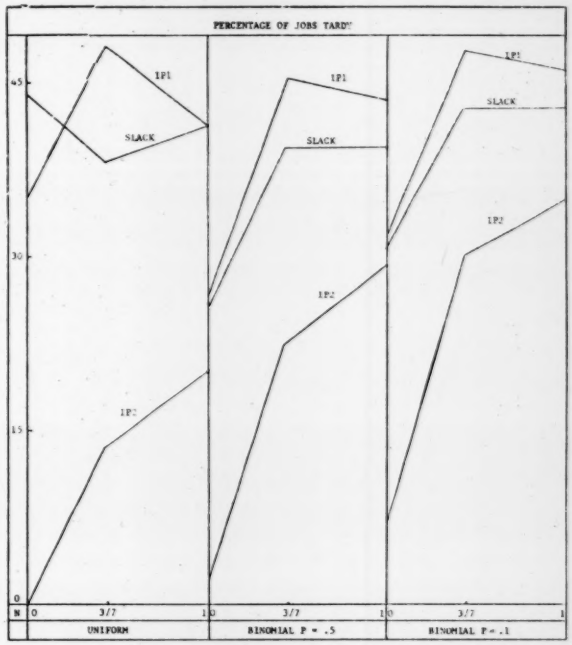


Fig. 2. Percentage of Jobs Tardy

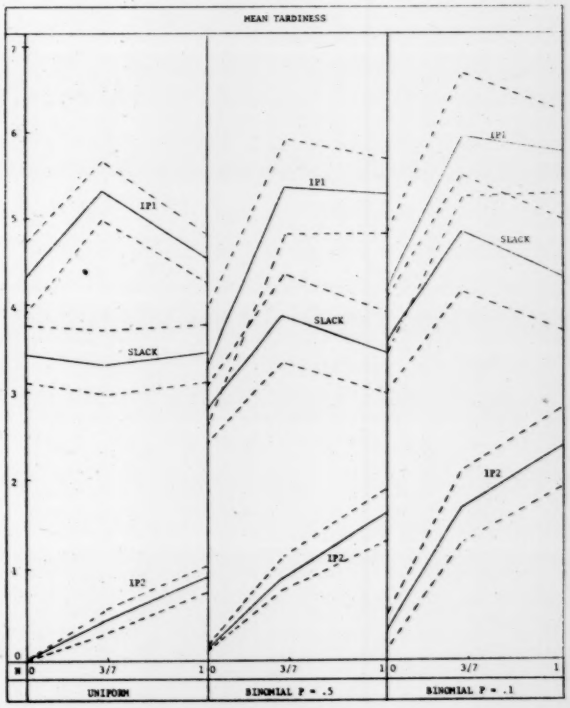


Fig. 3. Mean Tardiness Statistics

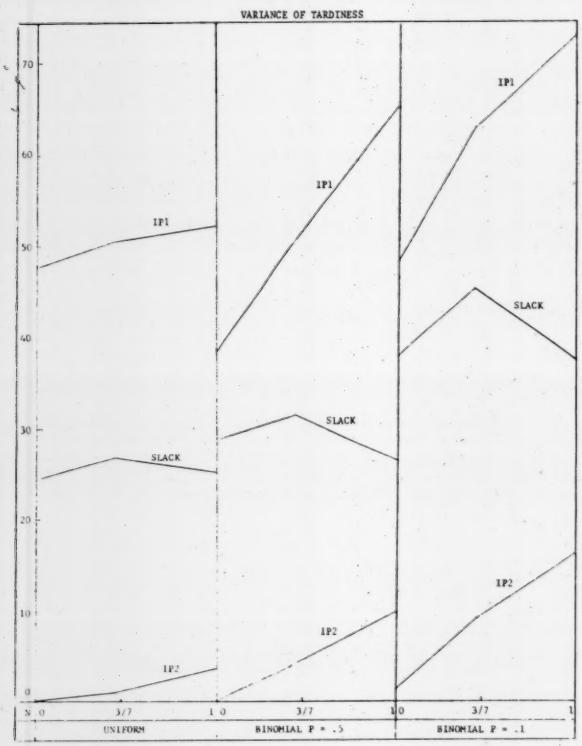


Fig. 4. Variance of Tardiness.

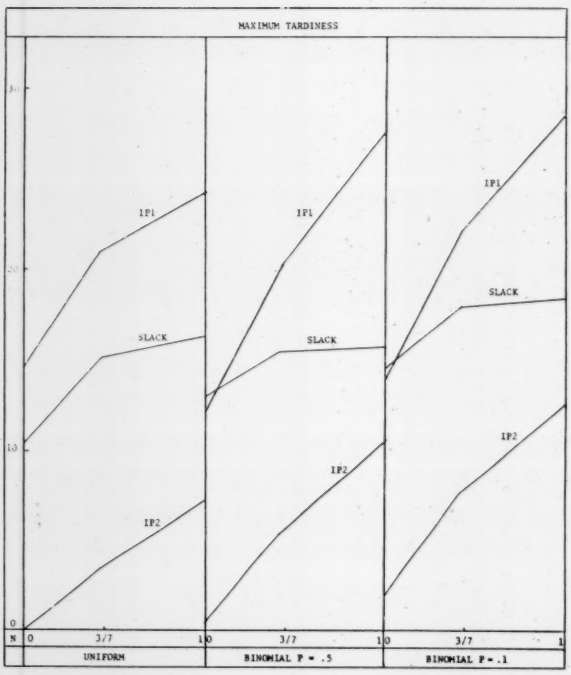


Fig. 5. Maximum Tardiness

The costs of data collection and computation can be divided into those relating to generation of the centralized schedule at review times and those relating to calculation of priorities on the shop floor. The computational times for generating the centralized schedule will depend upon the problem size and characteristics. The problems used in this paper required approximately 1 second per HSP run on an IBM 360-91. With an average of two iterations for the step sizes used in tightening due dates for HSP, the total computation time was approximately 2 seconds per problem. No general results are available but previous research (see [2]) indicates that, for up to 100 operations, the computing time in seconds per HSP iteration on an IBM 360-91 is approximately (number of operations/30). Two problems with 400 operations took 14 and 24 seconds of computer time each. Data required for each review is only that normally available in production control systems.

Priority calculation on the floor requires the centrally generated schedule for each machine, estimated arrival time, due-dates, estimate processing time on the given machine, and estimated remaining processing time on the job. All of this information is either required for the normal SLACK calculation or available from the output of centralized scheduling procedure. The number of calculations on the shop floor using IP2 can be fewer than with SLACK since the amount of slack need only be calculated for new jobs and for centrally scheduled operations when new jobs are available.

The results also provide some insights into the two implementation strategies. The nondelay strategy with preference given to centrally scheduled jobs (IP1) appears to be inferior to a strategy based on delays and integration of new jobs (IP2). Previous research [3] indicates that the relatively poor performance of IP1 compared to SLACK for $N = 0$ may be due to the characteristics of the problem used. However, we speculate that the increase in superiority of SLACK over IP1 for higher values of N reflects the lack of a means for integrating new jobs with those centrally scheduled. These findings and speculations indicate that continued research along the lines suggested in the next section may be valuable.

Summary and Indicated Extensions

Summary. Building upon the results of earlier research, a procedure is proposed for utilizing solutions to a job shop scheduling problem with due dates, static arrivals, and deterministic processing times in a dynamic environment with statistical processing times. Application of the procedure requires that a description of current shop status be available at points in time designated as review times. Two options are considered for implementing the priorities on the shop floor. The paper includes a detailed illustrative example and the results of experiments with an evaluative example. The principal conclusion of the research is that periodic centralized generation of schedules, implemented using option IP2, merits consideration as an effective sched-

uling procedure in dynamic job shop processes with due dates.

The two strategies for implementation evaluated could obviously be extended to consider other implementation procedures. For instance, the interrelationship between the delay and integration features combined in IP2 could be investigated as well as different methods for accomplishing each. In addition, more extensive evaluation of IP2 may be valuable.

Another area which appears fruitful for further research is the interactive use of the procedure. The iterative procedure used in this research was based on the single criterion, minimize maximum tardiness, and a simple search procedure. There is reason to believe that an individual faced with a scheduling problem may be interested in more than maximum tardiness. Moreover, an individual's intuition may prove superior to any prescribed iterative procedure such as used here for seeking improved solutions. This suggests the central thesis that in an interactive mode the machine is more useful as a source of information for problem identification than as a problem solver.

Guided by this thesis, the authors [4] have developed an interactive version of HSP. That program (ISP) includes provisions for modification of goals and information displays to aid the scheduler in applying ad hoc resource adjustments. ISP provides a mechanism for extending the use of the scheduling procedure described in this paper into the realm of interactive scheduling research.

Appendix A

The HSP procedure is designed to solve the following formulation of the static job shop problem.

Minimize [extent of precedence violations]
subject to:
1. All job due dates are satisfied
2. Fixed machine capacity

The basic structure of the multi-pass procedure is outlined below. For a more detailed description see [2].

Constraints

Because it is based upon the above formulation, the procedure schedules each machine subject to firm constraints imposed by the fixed machine capacity and job due dates.

(a) *Machine capacity constraints.* These constraints are imposed by taking the given processing time for each operation as fixed, by requiring that only one operation can be in process on a machine at any given time, and by not allowing any operation to be scheduled to begin before its earliest possible start time (EPST) given by the total processing time of the preceding operations on the job.

(b) *Due date constraints.* These constraints are imposed by not allowing any operation to be scheduled to begin after its latest possible start time (LPST) given by job due date minus remaining processing time for the job.

Criterion Function

The criterion function for the formulation is to minimize the extent of precedence violations among the operations in the schedule. The schedule requirements imposed on the operations for a given machine by precedence requirements and the current schedule on other machines are transmitted by computing an earliest start time (EST) and latest start time (LST) for each operation. These EST's and LST's change as the schedule changes and provide a means for measuring the value of the criterion function at any time. Operationally, the scheduling procedure tries to honor the EST's and LST's, but they are treated as soft constraints that can be violated to the extent necessary to schedule a machine. The essence of the scheduling procedure is a heuristic adjusting process designed to minimize the value of the criterion function as measured by the extent of the EST and LST violations.

Forces and Pseudo Constraints

The violation of an EST or LST indicates the need to reschedule an adjacent operation on the same job. When an operation is forced in this way, special attention is given to the need to reschedule the operation by using a set of pseudo constraints designated as EPST-P and LPST-P. The scheduling procedure treats these as intermediate in rigidity between the soft (EST and LST) and firm (EPST and LPST) constraints. These pseudo constraints provide the mechanisms for adjusting a machine's schedule to respond to forces that have resulted from the scheduling of other machines.

Classes of Forces

The scheduling procedure recognizes different classes of forces. A hierarchy of force classes ranging from most desirable to least desirable is established. When scheduling a machine, the operations in the class involving no forces are used first. A priority rule is used to select among the operations in that class. If all operations on the machine can be scheduled within that class, scheduling is begun on the next machine. If unscheduled operations remain, a choice must be made between (1) revising the current partial schedule on the machine in a renewed attempt to complete the scheduling of the machine without causing any forces; or (2) allowing operations to be scheduled from less desirable classes involving forces. The choice centers upon the question of how hard to try to schedule a machine without creating a force. A mode structure allows this question to be addressed operationally.

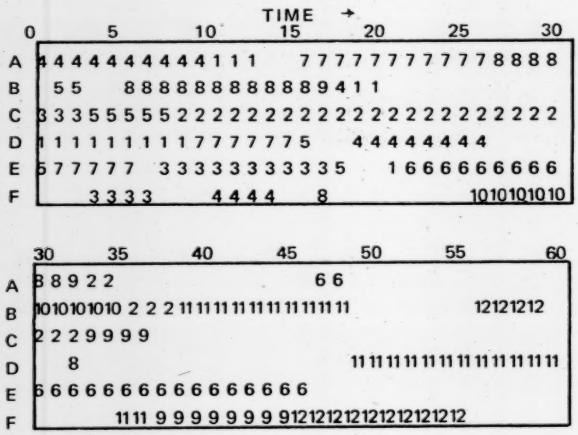
Scheduling Modes

A hierarchy of scheduling modes is established, paralleling the class structure. Each scheduling mode allows operations to be scheduled from a progressively larger subset of the force classes. Within each mode the preferred class structure is maintained. Before changing from one scheduling mode to the next, a back tracking routing can be used

to attempt to reschedule the machine within the current scheduling mode.

Appendix B

The Gantt chart for the illustrative example with IP2 priority implementation is given below. An example of the use of enforced idle time occurs at time $t = 6$ on machine E. At that time, job 6 is available for processing on the machine but IP2 calls for the machine to be held idle awaiting job 3, due to arrive from machine F.



Appendix C

The data for the *original jobs* in the experiments reported in Section 2 is given below:

Job	Arrival Time	Job Routing: Machine (Expected Processing Time)	Due Date
1	0	D(7) A(3) B(2) E(1)	26
2	0	C(20) A(2) B(2)	48
3	0	C(4) F(3) E(10)	34
4	0	A(10) F(3) B(1) D(8)	44
5	0	E(1) B(3) C(6) D(1) E(1)	24
6	0	E(22) A(3)	50
7	0	E(5) D(7) A(13)	50

The data for *new jobs* arriving during the scheduling period was generated as follows:

Poisson arrival process with mean arrival rate dependent on N.

Number of operations on job equals random integer from 1 to 6.

Each machine is equally likely for any operation (subject to the constraint that consecutive operations on a job be on different machines).

Expected processing time for each operation equals random integer between 1 and 10.

Job due date equal job arrival time plus two times total expected processing time of job.

References

[1] Conway, R. W., Maxwell, W. L., and Miller, L. W., *Theory of Scheduling*, Addison-Wesley (1967).

[2] Holloway, Charles A., and Nelson, Rosser T., "An Alternative Formulation of the Job Shop Problem with Due Dates," *Management Science*, **20** 1 (September 1973).

[3] Holloway, Charles A., and Nelson, Rosser T., "Job Shop Scheduling with Due Dates and Variable Processing Times," *Management Science*, **20** 9 (May 1974).

[4] Holloway, Charles A., and Nelson, Rosser T., "An Interactive Program for the Job Shop Scheduling Problem with Due Dates," Stanford University Graduate School of Business, Technical Report No. 22 (December 1972).

[5] Holloway, Charles A., and Nelson, Rosser T., "Job Shop Scheduling with Due Dates and Overtime Capability," *Management Science*, **21** 1 (September 1974).

[6] Holloway, Charles A., and Nelson, Rosser T., "Job Shop Scheduling with Due Dates and Operation Overlap Feasibility," *AIIE Transactions*, **7** 1 (March 1975).

[7] Peters, Michael H., "Optimal Static Job Shop Scheduling under Dynamic Conditions," PhD Dissertation, Indiana University (1971).

[8] Wong, Ruby Mei-Lun, "An Evaluation of Priority Implementation Procedures in a Job Shop Model with Due Dates, Dynamic Arrival Process, and Statistical Processing Times," Master's Thesis, UCLA (1973).

Ruby Mei-Lun Wong holds the BA and MS Degrees from UCLA. She is presently a PhD candidate in the Graduate School of Business, Stanford University. She was enrolled in the DBA Program at Harvard University and is a member of Phi Beta Kappa, Pi Mu Epsilon and Beta Gamma Sigma honorary fraternities.

Dr. Rosser T. Nelson is a Professor in the Graduate School of Management at UCLA. His current research interests are in job shop and resource constrained project scheduling, decomposition approaches in scheduling and man-machine interactive scheduling. Dr. Nelson received a BS Degree from Syracuse University and MS and PhD Degrees from UCLA. He is a member of TIMS and ORSA.

Dr. Charles A. Holloway is Associate Professor of Management Science in the Graduate School of Business, Stanford University. His current research is in job shop and large project scheduling, decomposition approaches in scheduling, mathematical programming with interactive identification and optimization, design of interactive graphic systems for managers and comparability of information systems. Dr. Holloway received the BSEE Degree from the University of California at Berkeley and MS and PhD Degrees from UCLA. He is a member of ORSA and TIMS.

TECHNICAL NOTE

Extensions of the Multi-Period Facility Phase-Out Model: New Procedures and Application to a Phase-In/Phase-Out Problem

GARY M. RODMAN
State University of New York at Binghamton
Binghamton, NY 13901

LEROY B. SCHWARZ
University of Rochester
Rochester, NY 14627

Abstract: This paper extends the multi-period facility phase-out model of Roodman and Schwarz in two ways: First, the branch-and-bound procedure is improved by the incorporation of two new lower bounds, which are both easier to obtain and more powerful than those originally proposed, and the addition of two new simplifications. Second, the model itself is generalized to solve a phase-in/phase-out problem. The improved branch-and-bound procedure is directly applicable to the static facility location models of Efroymson and Ray and Khumawala, and has consistently solved sample static problems faster than the methods proposed by these authors.

■ This paper extends the multi-period facility phase-out model of Roodman and Schwarz [3] in two ways:

1. The optimum finding branch-and-bound procedure is improved via the incorporation of improved techniques for computing lower bounds on the value of any given partial assignment (of open and closed facilities) *and* the addition of two new simplifications, which substantially reduce the number of partial assignments requiring investigation.

2. The facility phase-out model is generalized to solve a facility phase-in/phase-out problem.

In the following two sections we introduce the new lower bounds and the new simplifications. The next section develops a model for the phase-in/phase-out problem. The final section presents the results of computational tests on the originally proposed and improved procedures for a sample of phase-in, phase-out, and static problems.

In presenting the new procedures we shall assume that the reader is familiar with the terminology and notation of [3]. For clarity, references to equations in [3] will be preceded by "3:"; for example, Eq. (8) in [3] will be denoted "(3:8)".

Received March 1976; revised October 1976.

New Lower Bounds

In [3] the optimal solution to problem (3:1) - (3:7) is obtained by a branch-and-bound procedure in which lower bounds on the value of the best completion(s) of a given partial assignment A are computed by solving the trivial linear program resulting from elimination of constraints (3:6) and (3:7). This bounding procedure, called the LP bound, is essentially that proposed by Efroymson and Ray [1] for the static facility location problem. In this section, two new lower bounds are introduced which are simpler and also more powerful than the LP bound for both the phase-out (and corresponding phase-in) and the static problems.

Consider any partial assignment A. Let

U=the value (minimum cost) of the completion of A in which all free facilities are *closed* in all corresponding time periods; and

V=the value of the completion of A in which all free facilities are *opened* in all corresponding time periods.

U and V are upper bounds on the value of the best completion of A. As defined in (3:11), β_{it}, $i \in K_{2t}$, is an upper bound on the change in U (on the savings) associated with fixing facility i open in periods E_i through t. For partial

assignment A, define $K_2 = \cup_{t=1}^{T} K_{2t}$, the set of all facilities that are free in at least one time period. Consider the completion of A in which each facility i, $i \in K_2$, is open in some given period r_i and closed in all subsequent periods. A lower bound on the value of this particular completion is:

$$B_i' = U - \sum_{i \in K_2} \beta_{ir_i} . \tag{1}$$

Define

$$B_1 = U - \sum_{i \in K_2} \max_{E_i \leq r_i \leq L_i} \{\beta_{ir_i}\}. \tag{2}$$

B_1 minimizes (1) for all choices of r_i, $i \in K_2$, and is therefore a lower bound on the value of the best completion of A. If $\bar{z}$ denotes the value of the incumbent (best known feasible solution) and it is determined that $\bar{z} \leq B_1$, the study of A and its completions may be terminated.

By similar reasoning, define

$$B_2 = V - \sum_{i \in K_2} \max_{E_i \leq r_i \leq L_i} \{\gamma_{ir_i}\} \tag{3}$$

where γ_{it} is defined in (3:12). B_2 is also a lower bound on the value of the best completion of A, so that again if $\bar{z} \leq B_2$, the study of A may be terminated.

Note the simplicity of computing B_1 (B_2) given a partial assignment: for each free facility i merely pick the time period r_i which maximizes β_{it} (γ_{it}); subtract the sum of β_{ir_i}'s (γ_{ir_i}'s) from U (V) and the bound is at hand. Also note that when these new bounds replace the LP bound, Simplification 3 of [3] is no longer required since it is designed to improve the LP bound.

In the test problems below the bound used in the search was $\max\{B_1, B_2\}$. This bound is denoted the β–γ bound.

New Simplifications

The new lower bounds allow the construction of two new simplifications:

Simplification 4: For each $k \in K_2$ define

$$B_1^k = U - \sum_{\substack{i \in K_2 \\ i \neq k}} \max_{E_i \leq r_i \leq L_i} \{\beta_{ir_i}\}. \tag{4}$$

If $B_1^k \geq \bar{z}$, then completions of A better than the incumbent, if they exist, have $Y_{kt} = 1$ for $E_k \leq t \leq f$, where f is the smallest r for which $\beta_{kr} > B_1^k - \bar{z}$.

Proof: As defined, B_1^k is a lower bound on the value of the completions of A in which facility k is closed in all free periods. Thus, if $B_1^k \geq \bar{z}$, completions of A better than the incumbent, if they exist, must have facility k open in period E_k (and perhaps later). We may therefore augment A by fixing facility k open in E_k without eliminating completions of A that are better than the incumbent. Denote the augmented A by $\hat{A}$ and its corresponding B_1^k as:

$$\hat{B}_1^k = \hat{U} - \sum_{\substack{i \in K_2 \\ i \neq k}} \max_{E_i \leq r_i \leq L_i} \{\hat{\beta}_{ir_i}\}.$$

$\hat{U} = U - \beta_{kE_k}$, since, by definition, β_{kE_k} is the reduction in U from opening facility k in period E_k. Also, $\hat{\beta}_{ir} \leq \beta_{ir}$, $i \neq k$, because the β's decrease or remain unchanged as free facilities are opened. Thus,

$$\hat{B}_1^k \geq B_1^k - \beta_{kE_k} . \tag{5}$$

If $\hat{B}_1^k \geq \bar{z}$, then, by the above argument, $\hat{A}$ may itself be augmented by opening facility k in period $E_k + 1$. By (5), a sufficient condition for $\hat{B}_1^k \geq \bar{z}$ is $B_1^k - \beta_{kE_k} \geq \bar{z}$, or equivalently, $\beta_{kE_k} \leq B_1^k - \bar{z}$. Therefore if $\beta_{kE_k} \leq B_1^k - \bar{z}$, $\hat{A}$ may be augmented by opening facility k in period $E_k + 1$. This augmentation process continues period by period until facility k is opened in period f, where f is the smallest r for which $\beta_{kr} > B_1^k - \bar{z}$.

Simplification 5: For each $k \in K_2$ define

$$B_2^k = V - \sum_{\substack{i \in K_2 \\ i \neq k}} \max_{E_i \leq r_i \leq L_i} \{\gamma_{ir_i}\}. \tag{6}$$

If $B_2^k \geq \bar{z}$, then completions of A better than the incumbent, if they exist, have $Y_{kt} = 0$ for $f \leq t \leq L_k$, where f is the largest r for which $\gamma_{kr} > B_2^k - \bar{z}$.

Proof: The proof is similar to the proof for Simplification 4 except that the augmentation process moves backward through time from period L_k instead of forward from E_k.

The Phase-in/Phase-out Problem

Consider the following generalization of the facility phase-out problem: Given an initial configuration of existing facilities and forecasts of demand and costs over a T-period planning horizon, develop a plan for closing (phasing out) old facilities, opening (phasing in) new facilities, and modifying distribution schedules over the horizon that will minimize total discounted cost. A model for this *phase-in/phase-out* problem can be stated by making the following modifications in (3:1) through (3:7):

1. Assume that only the first R facilities ($R \leq M$), $i = 1, \ldots, R$, are open at the beginning of the planning horizon; the remainder, $i = R+1, \ldots, M$, are potential new sites.

2. Define $F_{it} = O_{it} - O_{it+1}$ for $i = R+1, \ldots, M$ and $t = 1, \ldots, T$, where O_{it} is the discounted fixed cost to open

facility i in period t and operate it in periods t through T, less the discounted salvage value of the facility at the beginning of the period $T+1$; $O_{iT+1} = 0$, $i = R+1, \ldots, M$. F_{it} may be viewed as the incremental fixed cost of opening facility i in period t rather than period $t+1$. The definition of F_{it} for $i = 1, \ldots, R$ is unchanged.

3. Replace constraints (3:7) with two sets of constraints,

$$Y_{it} \leqslant Y_{it-1} \quad i = 1, \ldots, R; \; t = 1, \ldots, T \tag{7a}$$

$$Y_{it} \geqslant Y_{it-1} \quad i = R+1, \ldots, M; t = 1, \ldots, T \tag{7b}$$

and set $Y_{i0} = 1, i = 1, \ldots, R$, and $Y_{i0} = 0, i = R+1, \ldots, M$.

The new constraints (7b) specify that a new facility may be opened in any time period, but once opened, must remain open through period T, just as existing facilities, once closed, must remain closed through period T. In other words, this model does not permit the same facility to be opened and closed (or closed and re-opened) during the planning horizon. This should pose no great problem in actual application. Note also that to model the pure phase-in problem, in which all facilities represent potential new sites, we may simply take $R = 0$.

This phase-in/phase-out model can be solved using an extension of the procedure previously described. For each phase-in facility $i, i = R+1, \ldots, M$, it is merely necessary to redefine the α_{it}, β_{it}, γ_{it} and δ_{it} of (3:10) through (3:13) as follows:

$$\alpha_{it} = \sum_{r=t}^{L_i} \Delta_{ir} \quad i = R+1, \ldots, M; \; t = E_i, \ldots, L_i \tag{8}$$

$$\beta_{it} = \sum_{r=t}^{L_i} \Omega_{ir} \quad i = R+1, \ldots, M; t = E_i, \ldots, L_i \tag{9}$$

$$\gamma_{it} = -\sum_{r=E_i}^{t} \Delta_{ir} \quad i = R+1, \ldots, M; t = E_i, \ldots, L_i \tag{10}$$

$$\delta_{it} = -\sum_{r=E_i}^{t} \Omega_{ir} \quad i = R+1, \ldots, M; t = E_i, \ldots, L_i. \tag{11}$$

Once these new definitions are introduced, lower bounds B_1 and B_2 can be computed and used to solve the phase-in/phase-out problem precisely as already described. The four simplifications, however, will be slightly different for phase-in and phase-out facilities. The differences are that whenever $i \leqslant R$, "opening" in Simplifications 1 and 4 occurs from critical period f *forward* to L_i rather than backward to E_i; and similarly, "closing" in Simplifications 2 and 5 moves *backward* from period f to E_i, rather than forward to L_i. For example, Simplification 1 assumes the following form:

For each $i = 1, \ldots, M$, determine

$$\alpha_{if} = \max_{E_i \leqslant f \leqslant L_i} \{\alpha_{it}\}.$$

If $\alpha_{if} \geqslant 0$ and $i \leqslant R$ ($i \geqslant R+1$), then there exists a best completion of A such that $Y_{it} = 1$ for $E_i \leqslant t \leqslant f$ (for $f \leqslant t \leqslant L_i$).

The heuristics presented in [3] are also applicable to the phase-in/phase-out problems, given the appropriate definitions for the β's and γ's.

Computational Tests

In order to test the efficiency of the new simplifications and lower bounds we solved 10 problems: 4 phase-out problems, 4 phase-in problems, and 2 static problems. Problems 1, 2, 3, and 4 are problems 4, 5, 7, and 8 of [3], respectively. Problems 5-8 are phase-in problems; problems 9 and 10 are static problems. Detailed parameters for these problems are available on request. Note that problem size varies from 2^{15} feasible solutions to 9^{15} feasible solutions $[(T+1)^M]$. We solved each problem using three different procedures: *Procedure 1*, using the LP bound and Simplifications 1 through 3 of [3]; *Procedure 2*, using the new β-γ bound and only Simplifications 1 and 2; *Procedure 3*, using the β-γ bound and Simplifications 1, 2, 4, and 5. Table 1 presents the results of our tests. All results were obtained on a Honeywell G635 computer.

In Section (1) of the table the values of M, N, and T are given for each test problem; and Section (2) presents the number of facilities which are open in period T of the optimal solution. Sections (3), (4), and (5) compare the performance of the branching rules and the different procedures. Section (3) indicates the number of partial assignments examined (called enumerations) by each rule in each procedure to obtain the optimal solution as the incumbent. Section (4) indicates the number of enumerations required to obtain and confirm the optimum with each procedure; Section (5) presents the run times in seconds for each branching rule using each procedure. Procedure 1 results appear in parentheses; Procedure 2 results in brackets.

In order to objectively test the branching rules and procedures, all searches began without initial incumbent solutions. (All computation times reported would improve, perhaps substantially, if the searches began with good incumbents such as are generated by the heuristics described in [3].) Missing figures indicate that the search using the given rule was terminated before completion. Termination occurred whenever 1000 seconds of computation time had elapsed or the number of unfathomed partial assignments reached 700.

If we examine first the test results for the four phase-out problems of [3], problems 1-4, for the largest β branching rule, it is clear that separately and in combination the β-γ bound and the new simplifications yield dramatic improvements over the original computational procedure. The run

Table 1

Description		Test Problem 1	2	3	4	5	6	7	8	9	10
(1) Problem Parameters	M	12	12	8	8	12	12	15	15	15	25
	N	40	50	40	50	30	30	30	30	30	50
	T	5	5	8	8	5	8	8	6	1	1
(2) No. of facilities open in period T of the optimal solution		3	3	0	0	3	6	5	5	6	6
(3) Number of Enumerations to reach optimum	β	(151) [161] 147	(287) [223] 219	(227) [187] 185	(95) [77] 37	(137) [227] 161	(279) [397] 127	(34) [15] 15	(14) [21] 21	(23) [11] 11	(590) [193] 193
	γ	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 626	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(22) [16] 16	(•) [730] 726
	α	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 471	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(54) [29] 29	(848) [647] 647
(4) Total Number of Enumerations	β	(525) [395] 201	(865) [639] 331	(945) [425] 299	(187) [91] 43	(303) [337] 189	(719) [585] 133	(488) [443] 73	(184) [219] 49	(49) [37] 21	(1132) [711] 371
	γ	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 877	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(78) [29] 25	(•) [1225] 1053
	α	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 473	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(92) [67] 51	(•) • 1169
(5) Run Time (seconds)	β	(314.2) [242.5] 160.6	(654.3) [462.4] 302.0	(521.7) [268.9] 211.8	(175.4) [92.0] 61.9	(192.8) [210.3] 141.4	(587.1) [479.3] 173.3	(721.7) [548.2] 169.2	(214.7) [220.1] 85.5	(16.5) [11.6] 8.9	(806.4) [472.6] 249.8
	γ	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 844.5	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(28.1) [11.2] 10.4	(•) [889.3] 794.7
	α	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] 719.9	(•) [•] •	(•) [•] •	(•) [•] •	(•) [•] •	(25.6) [17.3] 14.5	(•) [•] 755.3

time using Procedure 2 (in brackets) averages approximately two-thirds the run time of Procedure 1 (in parentheses), indicating the superiority of the β-γ bound over the LP bound. The run time of Procedure 3, using the new Simplifications, is again approximately two-thirds the run time of Procedure 2. Note that improvements in the run times are derived mostly from a reduction in the total number of enumerations required to reach and confirm the optimal solution (Section 4), although there is also some reduction in the number of enumerations to reach the optimal solution (Section 3). The results and interpretations for the four phase-in problems are similar.

The test results for Problems 9 and 10 indicate the superiority of Procedures 2 and 3 over those of [1] and [2] for solving the static problem. In order to see this note that for the static problem $\beta_i = \Omega_i$, $\alpha_i = \Delta_i$, and $\gamma_i = -\Delta_i$ (the t subscript has been omitted for clarity). Thus, in Procedure 1 we can replicate the procedure of [1] and [2] with the largest Ω, largest Δ, and smallest Δ branching rules.

For the β branching rule the β-γ bound reduces the run time of [2] by about one-third; the new simplifications provide additional improvement. For the γ and α branching rules only Procedure 3 reaches and confirms the optimal solution before termination.

Finally, note that, as in [3], the β branching rule continues to dominate the γ and α rules, regardless of the bounds or simplifications used.

References

[1] Efroymson, M. A. and Ray, T. L., "A Branch-Bound Algorithm for Plant Location," *Operations Research*, **14** 3, pp. 361-368 (May-June, 1966).

[2] Khumawala, B. M., "An Efficient Branch-Bound Algorithm for the Warehouse Location Problem," *Management Science*, **18** 12, pp. 718-732 (August 1972).

[3] Roodman, G. M. and Schwarz, L. B., "Optimal and Heuristic Facility Phase-out Strategies," *AIIE Transactions*, **7** 2, pp. 177-184 (July 1975).

Dr. Gary M. Roodman is an Associate Professor of Management at the State University of New York at Binghamton. He earned MBA and DBA degrees at the Graduate School of Business, Indiana University. His research interests include integer programming methods, facility location problems, and management of health care delivery. Professor Roodman is a member of TIMS, ORSA, and AIDS.

Dr. Leroy B. Schwarz is an Assistant Professor of Management at the Graduate School of Management, University of Rochester. He received his MBA and PhD from the Graduate School of Business, University of Chicago. Research interests include the management of network inventory systems, facility location problems, and the design and scheduling of automatic warehousing systems. Professor Schwarz is a member of TIMS and a member of ORSA.

TECHNICAL NOTE

An Order-Level Lot-Size Inventory Model for Deteriorating Items

Y. K. SHAH
School of Sciences
Gujarat University
Ahmedabad-380009 India

Abstract: A deterministic inventory model is developed for deteriorating items for a system that allows backlogging under the condition of immediate replenishments. The model is developed for any well-behaved probability distribution for the time to deterioration of an item. From the model developed, the earlier models developed by Ghare and Schrader and Covert and Philip can be obtained as particular cases by taking a suitable probability distribution for the time to deterioration of an item when shortage cost becomes large.

■ Recently there have been a few efforts of analyzing mathematical models of inventory for items with a stochastic lifetime, so that the items may deteriorate while in inventory. Ghare and Schrader [3] have developed an EOQ model wherein they have assumed the negative exponential distribution for the time to deterioration of an item so that, in their case, the instantaneous deterioration rate is constant. Covert and Philip [1] have obtained an EOQ model for items with a variable rate of deterioration by assuming the two-parameter Weibull distribution for the time to deterioration of an item. Misra [4] has developed a deterministic model for both a constant and a variable rate of deterioration of items for a system with a finite production rate. In all these models backorders or shortages are not permissible and the items start deteriorating from the instant they arrive into the inventory system. A more general approach would be one which permits already deteriorated items to be received by an inventory system as well as items which may start deteriorating in future. Philip [6] has developed such a model by assuming the three parameter Weibull distribution for the time to deterioration of an item.

In this note, an order-level lot-size inventory model is developed for items that deteriorate continuously in time in accordance with some well-behaved probability distribution $g(t)$ for the lifetime of an item. The system permits stockouts, and assumes instantaneous delivery and an infinite replenishment rate. By a well-behaved probability distribution for the time to deterioration of an item we shall mean a probability distribution for which the age-specific failure-rate function $\phi(t)$ (defined below) exists, is an explicit function of t, and is integrable with respect to t. From the general model developed here, when shortage cost approaches infinity, the results of Ghare and Schrader [3] and Covert and Philip [1] can be obtained as particular cases by taking a suitable probability distribution for the time to deterioration of an item.

Received March 1976; revised October 1976.

The Mathematical Model for A General Distribution of Deterioration

The model is developed on the following assumptions:

(i) The demand rate of R units per time unit is known and constant.

(ii) The replenishment rate is infinite; the replenishment size is constant. The lot size is q units per replenishment.

(iii) Shortages, if any, are made up in the beginning of a cycle.

(iv) There is no repair or replacement of deteriorated inventory during a given cycle.

(v) The inventory holding cost C_1 per unit per time unit, the shortage cost C_2 per unit per time unit, the replenishment cost C_3 per cycle and the unit cost C are known and constant during the period under considerations.

(vi) The distribution of the time to deterioration of an item has probability density function (p.d.f.) $g(t)$ ($t \geq 0$) and cumulative distribution function (c.d.f.) $G(t)$; so that, following Cox [2], the instantaneous or age-specific failure-rate function of an item is

$$\phi(t) = g(t) / \{1 - G(t)\}, \quad t \geq 0. \tag{1}$$

Note that $\phi(t)dt$ shows the probability of failure of an item during the infinitesimal period $(t, t+dt)$ given that it has not failed prior to t.

Let T denote the period of a cycle, q the lot-size and S the order-level. We analyze one cycle. At $t = 0$, an ordered batch of q units enter the inventory system, from which $(q-S)$ units are delivered towards backorders, leaving a balance of S units to be carried as initial inventory. Then the inventory level gradually decreases due to demands at the rate R per time unit and deterioration of items in inventory in accordance with the instantaneous rate function $\phi(t)$ up to the time $t = t_1$ (say). At $t = t_1$ the inventory level becomes zero. Further demands during (t_1, T) are backlogged. Let $t_2 = T - t_1$, then incidently we have

$$q = S + Rt_2 = S + RT - Rt_1 . \tag{2}$$

Let $Q(t)$ denote the inventory level of the system at time t $(0 \leq t \leq T)$, then $Q(t)$ is a function of the demand rate R and the deterioration function $\phi(t)$. Noting that during (t_1, T) no units are carried in inventory, the differential equation that describes the instantaneous states of $Q(t)$ over the period $(0, T)$ is given by

$$\begin{aligned} \frac{dQ(t)}{dt} + \phi(t)Q(t) &= -R, \quad 0 \leq t \leq t_1 \\ \frac{dQ(t)}{dt} &= -R, \quad t_1 \leq t \leq T. \end{aligned} \tag{3}$$

The solution of (3), after adjusting for the constants of integration, is

$$\begin{aligned} Q(t) &= \frac{R}{u(t)} [U(t_1) - U(t)], \quad 0 \leq t \leq t_1 \\ &= R(t_1 - t), \quad t_1 \leq t \leq T. \end{aligned} \tag{4}$$

where

$$u(t) = \exp\left[\int_0^t \phi(t)\,dt\right], \quad t \geq 0 \tag{5}$$

and

$$U(t) = \int_0^t u(t)dt, \quad t \geq 0. \tag{6}$$

Note that the initial inventory level is

$$Q(0) = S = RU(t_1) \tag{7}$$

and hence, from (2), the lot-size is

$$q = R\,[U(t_1) + T - t_1]\,. \tag{8}$$

We find that the two decision variables q and S are explicit functions of t_1 and T. We then express the cost function of the system in terms of t_1 and T, which is

$$\begin{aligned} K(t_1, T) &= \frac{C}{T}(S - Rt_1) + \frac{C_1}{2T} St_1 + \frac{C_2 R}{2T}(T - t_1)^2 + \frac{C_3}{T} \\ &= \frac{CR}{T}[U(t_1) - t_1] + \frac{C_1 R}{2T} t_1 U(t_1) + \frac{C_2 R}{2T}(T - t_1)^2 \\ &\quad + \frac{C_3}{T}. \end{aligned} \tag{9}$$

In obtaining the cost function (9) it is conveniently assumed that the function $Q(t)$ is approximately linear in t over $(0, T)$ so that the average number of units carried in inventory per unit of time is

$$\tfrac{1}{2}[Q(0) + Q(t_1)] \cdot t_1/T = \frac{1}{2T} St_1 .$$

To find the optimum values of t_1 and T, $\partial K/\partial t_1 = 0$ and $\partial K/\partial T = 0$ give

$$2C\{u(t_1) - 1\} + C_1\{t_1 u(t_1) + U(t_1)\} + 2C_2 t_1 = 2C_2 T \tag{10}$$

and

$$\left[2C\{U(t_1) - t_1\} + C_1 t_1 U(t_1) + \frac{2C_3}{R}\right] + C_2 t_1^2 = C_2 T^2 \tag{11}$$

respectively. Eliminating T from (10) and (11) we get

$$\begin{aligned} &[2C\{u(t_1) - 1\} + C_1\{t_1 u(t_1) + U(t_1)\}]^2 \\ &+ 8CC_2\{t_1 u(t_1) - U(t_1)\} + 4C_1 C_2 t_1^2 u(t_1) - \frac{8C_2 C_3}{R} = 0. \end{aligned} \tag{12}$$

Solving (12) we can obtain the optimum value of $t_1 = t_{1o}$ (say), and substituting it into (10) we get the optimum cycle time $T = T_o$ (say). Finally, substituting $(t_1, T) = (t_{1o}, T_o)$, (9) will give the average minimum cost $K(t_{1o}, T_o)$ per unit of time, (8) will give the optimum lot-size $q = q_o$ and (7) will give the optimum order-level $S = S_o$. Generally, numerical methods will be required to solve (12).

If shortages are not allowed i.e. if we take $C_2 \equiv \infty$, (10) gives $t_1 = T$. In this case $q = S$ and the cost function (9) becomes

$$K(T) = \frac{CR}{T} U(T) - CR + \tfrac{1}{2} C_1 R\, U(T) + \frac{C_3}{T}. \tag{13}$$

and for the optimum value of T the equation to be solved is

$$C[Tu(T) - U(T)] + \tfrac{1}{2} C_1 T^2 u(T) - \frac{C_3}{R} = 0 . \qquad (14)$$

Eqn (14) can also be obtained if we divide (12) throughout by C_2 and take the limit as $C_2 \to \infty$.

If $\phi(t) = 0$ there is no deterioration and $u(t) = 1$, $U(t) = t$ for all t. In this case the cost function (9) becomes

$$K(t_1, T) = \frac{C_1 R t_1^2}{2T} + \frac{C_2 R (T - t_1)^2}{2T} + \frac{C_3}{T} \qquad (15)$$

and then, from (12) and (10), the optimum values of t_1 and T are

$$t_1^{(o)} = \left[\frac{2C_3}{C_1 R} \quad \frac{C_2}{C_1 + C_2}\right]^{\frac{1}{2}} \qquad (16)$$

and

$$T^{(o)} = \left[\frac{2C_3}{C_1 R} \quad \frac{C_1 + C_2}{C_2}\right]^{\frac{1}{2}} \qquad (17)$$

These are the standard formulas for obtaining (t_1, T) in the similar model for non-deteriorating items, as given by Naddor [5, p. 82].

From the general model developed above two particular cases are developed by taking the negative exponential distribution and the Weibull distribution for the time to deterioration of an item.

Negative Exponential Distribution for the Time to Deterioration of an Item

Let the p.d.f. of the time to deterioration of an item be

$$\begin{aligned} g(t) &= \theta \exp(-\theta t) \quad , \quad t \geq 0, \theta > 0 \\ &= 0 \quad , \quad \text{otherwise} \end{aligned} \qquad (18)$$

with its c.d.f. given by

$$G(t) = 1 - \exp(-\theta t) \qquad t \geq 0 \qquad (19)$$

where $\theta^{-1} = E(t) = \int_0^\infty t\, g(t)\, dt$ is the mean lifetime of an item. Using (18) and (19) in (1) the instantaneous failure rate function is

$$\phi(t) = \theta \qquad (20)$$

and then (5) and (6) give

$$u(t) = \exp(\theta t) \quad , \quad 0 \leq t \leq t_1 \qquad (21)$$

and

$$U(t) = \theta^{-1} \{\exp(\theta t) - 1\}, \; 0 \leq t \leq t_1 . \qquad (22)$$

In this case the cost function (9) becomes

$$K(t_1, T) = \frac{CR}{\theta T}\{\exp(\theta t_1) - 1\} - \frac{CRt_1}{T} + \frac{C_1 R t_1}{2\theta T}\{\exp(\theta t_1) - 1\} + \frac{C_2 R}{2T}(T - t_1)^2 + \frac{C_3}{T} \qquad (23)$$

and Equation (12) becomes

$$\left[2C\{\exp(\theta t_1) - 1\} + C_1\left(\frac{1}{\theta} + t_1\right)\exp(\theta t_1) - \frac{C_1}{\theta}\right]^2 \qquad (24)$$

$$- 8CC_2\left[\left(\frac{1}{\theta} - t_1\right)\exp(\theta t_1) - \frac{1}{\theta}\right] + 4C_1 C_2 t_1^2 \exp(\theta t_1) - \frac{8C_2 C_3}{R} = 0 .,$$

Solving (24) we get $t_1 = t_{10}$ (say) and substituting it in

$$2C\{\exp(\theta t_1) - 1\} + C_1\left\{\left(\frac{1}{\theta} + t_1\right)\exp(\theta t_1) - \frac{1}{\theta}\right\} + 2C_2 t_1 = 2C_2 T \qquad (25)$$

the optimum cycle time $T = T_o$ can be obtained.

Finally, for the optimum policy, the optimum values of the order-level S and the lot-size q are respectively

$$S = S_0 = R\, U(t_{1o}) = \frac{R}{\theta}\{\exp(\theta t_{1o}) - 1\} \qquad (26)$$

and

$$q = q_o = R\left[T_o - t_{1o} + \frac{1}{\theta}\{\exp(\theta t_{1o}) - 1\}\right] \qquad (27)$$

Using the series form of the exponentials and ignoring terms with second and higher order powers of θ under the assumption that $\theta \ll t_1$, Eqns. (23) to (27) become

$$K(t_1, T) = \frac{CR\theta t_1^2}{2T} + \frac{C_1 R t_1^2}{2T}(1 + \tfrac{1}{2}\theta t_1) + \frac{C_2 R}{2T}(T - t_1)^2 + \frac{C_3}{T} , \qquad (28)$$

$$C_1(3C_1 + 2C_2)\theta t_1^3 + 2(C_1^2 + C_1 C_2 + 2CC_1\theta + CC_2\theta) t_1^2 - \frac{4C_2 C_3}{R} = 0, \qquad (29)$$

$$C\theta t_1 + C_1 t_1 (1 + \tfrac{3}{4}\theta t_1) + C_2 t_1 = C_2 T . \qquad (30)$$

$$S = S_o = R\, t_{1o}(1 + \tfrac{1}{2}\theta t_{1o}) \qquad (31)$$

and

$$q = q_o = RT_o + \tfrac{1}{2} R\theta t_{1o}^2 \qquad (32)$$

respectively.

When $\theta = 0$ there is no deterioration and all the above equations reduce to those of the corresponding model for nondeteriorating items.

When $C_2 \equiv \alpha$, $t_1 = T$ and Eqn. (29) becomes

$$\frac{C\theta R}{2} + \frac{C_1 \theta T}{2} + \frac{C_1 R}{2} - \frac{C_3}{T^2} = 0$$

which is the same as that given by Ghare and Schrader [3].

Weibull Distribution for the Time to Deterioration of an Item

Assume now that the p.d.f. of the time to deterioration of an item is

$$g(t) = \alpha\beta\, t^{\beta-1} \exp(-\alpha t^{\beta}) \quad , t \geq 0,\ \alpha > 0,\ \beta > 0 \qquad (33)$$

$$= 0 \quad , \text{otherwise}$$

where α is the scale parameter and β is the shape parameter of the distribution. For this distribution the function $\phi(t)$, $u(t)$ and $U(t)$ are

$$\phi(t) = \alpha\beta\, t^{\beta-1} \qquad (34)$$

$$u(t) = \exp(\alpha t^{\beta}) \qquad (35)$$

and

$$U(t) = \int_0^t u(t)\, dt = \sum_{n=0}^{\infty} \frac{\alpha^n t^{n\beta+1}}{n!\,(n\beta+1)} = W(\alpha, \beta, t) \text{(say)} \qquad (36)$$

Then, in the present case, the cost function (9) becomes

$$K(t_1, T) = \frac{R}{T}\left(C + \frac{C_1 t_1}{2}\right) W(\alpha, \beta, t_1) - \frac{CRt_1}{T} \qquad (37)$$

$$+ \frac{C_2 R}{2T}(T - t_1)^2 + \frac{C_3}{T}$$

and for the optimum value of t_1 the pertinent equation is

$$[(2C + C_1 t_1)\exp(\alpha t_1^{\beta}) + C_1 W(\alpha, \beta, t_1) - 2C]^2 \qquad (38)$$

$$+ 8CC_2 [t_1 \exp(\alpha t_1^{\beta}) - W(\alpha, \beta, t_1)]$$

$$+ 4C_1 C_2 t_1^2 \exp(\alpha t_1^{\beta}) - \frac{8C_2 C_3}{R} = 0.$$

If $t_1 = t_{1o}$ is the optimum value of t_1 obtained by solving (38) then the optimum cycle time can be obtained by substituting it into

$$(2C + C_1 t_1)\exp(\alpha t_1^{\beta}) + C_1 W(\alpha, \beta, t_1) - 2C + 2C_2 t_1 = 2C_2 T \qquad (39)$$

The optimum order-level and lot-size are respectively

$$S = S_o = R\, U(t_{1o}) = R\, W(\alpha, \beta, t_{1o}) \qquad (40)$$

and

$$q = q_o = R\,[T_o - t_{1o} + W(\alpha, \beta, t_{1o})]\ . \qquad (41)$$

If shortages are not allowed then taking $C_2 \equiv \alpha$ we have $t_1 = T$ and the cost function (37) becomes

$$K(T) = \frac{R}{T}(C + \tfrac{1}{2}C_1 T)\, W(\alpha, \beta, T) - CR + \frac{C_3}{T}\ . \qquad (42)$$

For obtaining the optimum T, the respective equation will be

$$RT^2\left(\frac{C}{T} + \frac{C_1}{2}\right)\exp(\alpha T^{\beta}) - CR\, W(\alpha, \beta, T) - C_3 = 0. \qquad (43)$$

Equations (42) and (43) are the same as those given by Covert and Philip [1].

A Numerical Example

To illustrate the model, a hypothetical system was assumed with the following constants:

C = \$50.00 per unit, C_1 = \$ 0.02 per unit/day,

C_2 = \$ 0.10 per unit/day, C_3 = \$25.00 per order

R = 10 units per day.

The time to deterioration of an item was arbitrarily assumed to have an exponential distribution with mean life θ^{-1} = 500 days, i.e. $\theta = 0.002$. Eq. (24) was solved by using the Newton-Raphson iterative procedure and the optimum values of t_1 and T were found to be

$$t_{1o} = 4.3368311 \text{ days}$$

$$T_o = 9.5605914 \text{ days}.$$

The optimal values of the order-level S and lot-size q for the problem are respectively

$$S_o = 43.5565 \text{ units}$$

and

$$q_o = 95.7941 \text{ units}.$$

The minimum total cost per time unit is $K(t_{1o}, T_o)$ = \$5.2237 per day.

To check the optimality of the solution obtained, the total variable cost was computed for a few other values of (t_1, T). They are shown below.

(t_1, T) (days)	$K(t, T)$ (in \$)
(4, 9)	5.2340
(5, 10)	5.2538
(4, 10)	5.2606
(5, 11)	5.2761)

Acknowledgments

I am thankful to Dr. M. C. Jaiswal for his guidance and helpful discussions.

References

[1] Covert, R. P. and Philip, G. C., "An EOQ Model for Items with Weibull Distribution Deterioration," *AIIE Transactions*, **5**, 4, pp. 323-326 (1973).

[2] Ghare, P. M. and Schrader, G. F., "A Model for Exponentially Decaying Inventory," *Jr. of Ind. Engg.*, **XIV**, 243, (1963).

[3] Cox D. R., *Renewal Theory*, John Wiley and Sons, Inc. (1963).

[4] Misra, R. B., "Optimum Production Lot-size Model for a System with Deteriorating Inventory," *Int. Jr. of Prod. Res.*, **13** 495-505 (1975).

[5] Naddor, E., *Inventory Systems*, John Wiley and Sons, Inc. (1966).

[6] Philip, G. C., "A Generalized EOQ Model for Items with Weibull Distribution Deterioration, *AIIE Transactions*, **6**, 2, 159-162 (1974).

Professor Yashvant Shah is a Lecturer in Statistics at Gujarat University. He holds the BSc and MSc degrees from Gujarat University and is a life member of Gujarat Statistical Association. He is currently conducting research in operations research and inventory control.

Book Review

Queueing Systems. Volume I. Theory

Leonard Kleinrock

John Wiley & Sons, Inc., 605 Third Ave., New York, N.Y. 10016. (417 pages; $19.95) 1975.

Written as the first of the two-volume series, this book provides a comprehensive treatment of Queueing Theory. Industrial Engineers will hardly notice that the manuscript evolved from lecture notes used in a one-year course in Computer Science at the University of California, Los Angeles. True to its subtitle, "Theory," there are very few references to applications. Industrial Engineers who are seeking case studies and real-life examples may find that the second volume on "Applications" is too much geared toward "computer networks" and is not directly applicable to most IE problems.

Volume I is divided into three parts, Elementary Queueing Theory containing two chapters, Intermediate Queueing Theory containing three chapters, and a one-chapter Advanced Materials followed by two appendices on transforms and probability theory.

The strength of this book lies in its refreshing approach to the taxonomy of Queueing Theory models. Though the traditional three-part descriptor (A/B/m) is used, the emphasis is placed upon classifications of stochastic processes rather than upon queueing models. Thus, in treating a type of stochastic process, Kleinrock is free to switch from traditional models to more innovative models. Typical of this taxonomy is a delightful Venn diagram in Chapter Two that graphically illustrates the relationships among Semi-Markov Process, Markov Process, Random Walk, Renewal Process, Birth-Death Process, and Poisson Process. Because the author has taken the time to have a bird's-eye view of all stochastic processes before delving into stereotype queueing models, he is able to expand the scope of his "Queueing Theory" to encompass situations that are beyond the domain of the traditional texts. As any practicing Industrial Engineer knows, a textbook model seldom applies to a practical situation. The broader scope of the theory is likely to have resulted from Kleinrock's practical experience and attempt to describe all of the theory necessary in analyzing computer network systems discussed in Volume II.

Two consequences of this approach are inevitable. First, the author had to rely heavily upon Markov chains and transform theories, both the discrete z-transforms and the continuous Laplace transforms, in order to provide enough flexibility in treating non-traditional situations. Second, the reader must be knowledgeable not only in Probability Theory, as Kleinrock acknowledges, but also in traditional queueing models and broad area of stochastic processes before he can fully appreciate the significance of this book's contribution. For example, the afore-mentioned Venn diagram in Chapter Two will not be appreciated by an individual trying to learn Queueing Theory for the first time; in fact, it will appear as an unnecessary complication.

Kleinrock attempts to relieve the monotony of academic discussion by trying to be light-hearted wherever possible. Personal remarks and exclamation marks are abundant (e.g., "We never claimed it would be simple!") and the author uses a one-page synopsis preceding each Section to effectively warn his readers of what lies ahead (e.g., "Not even the average time is known for this queue!") And if you ever wonder where the author has come up with names like Abra, Zeus, Oriac, and Sucsamad in his examples, it helps to remember that the manuscript must have been prepared at the height of the Arab oil crisis when even a "backward" solution to the problem was appreciated.

Ample acknowledgment is given to other works. Conspicuous by absence, however, is the name of S. M. Ross. Many Markov chain models in Kleinrock's book resemble those treated by Ross in his earlier books of 1970 and 1972. In fact, even the approaches are similar. It's hard to imagine that a UCLA professor has not heard of the work by a colleague in the IEOR Department at the U.C. of Berkeley.

Kleinrock's book, Volume I, is strongly recommended to Industrial Engineers who are seriously interested in teaching or doing advanced work in stochastic processes. It will be a challenging text for a graduate course in Stochastic Processes, especially if the professor is looking for new ideas. The Summary of Important Results, at the end of the book, by itself, makes the book worth purchasing by a PhD student preparing for a comprehensive examination.

This book review is contributed by the Operations Research Division of AIIE.

Michael S. Inoue
Professor
Department of Industrial and General Engineering
Oregon State University
Corvallis, Oregon 97331

Engineering Economic Analysis

Donald G. Newnan

Engineering Press, San Jose, California. (469 pages; $12.95) 1976.

The author's basic objective in this book is to teach fundamental concepts of engineering economy *to engineers.* By limiting the intended audience in this manner, Dr. Newnan has attempted to provide a concise and relevant presentation of his subject matter.

This book is ideal for a one semester course at the undergraduate level, and its readability and numerous example problems make the book suitable for engineers and managers who desire to learn engineering economy on their own. Concerning classroom use of the book, the instructor will find helpful a manual written by the author that provides a typical course outline and an assortment of examination questions. Also, a comprehensive solutions manual for problems included at the end of each chapter is available.

After the concept of equivalence is discussed, the most popular methods of discounted cash flow analysis (i.e., present worth, uniform annual cost and rate of return) are described and illustrated in separate chapters. The comparison of alternatives on the basis of cash flow differences is described in a chapter to itself so that this important concept is drawn to the attention of the reader. In addition to traditional topics such as after-tax cash flow analysis, replacement analysis and economic analysis in government, it is noteworthy that Dr. Newnan has devoted a chapter to each of these subjects: inflation and deflation, estimation of future events (including simulation), capital budgeting and digital computer applications.

The basic shortcoming of this book is its omission of a brief discussion regarding elementary accounting concepts. However, this subject is frequently taught in other courses and can be easily reviewed by referring to an accounting text. Furthermore, the question of how to obtain data needed for engineering economy studies is generally avoided throughout the book. In many engineering economy texts, this matter is discussed in connection with a chapter devoted to basic cost accounting.

On balance *Engineering Economic Analysis* offers a refreshing, lucid treatment of principles basic to engineering economy. The many example problems, along with the absence of conspicuous typographical errors, make this book very readable and easy to understand. The author's straightforward explanation of the subject matter and his reasonably complete coverage of topics should place this book in the ranks of the better selling texts on engineering economy.

This book review was contributed by the Engineering Economy Division of AIIE.

William G. Sullivan
Department of Industrial Engineering
University of Tennessee
Knoxville, Tennessee

U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION: AIIE TRANSACTIONS Industrial Engineering Research and Development	2. DATE OF FILING: 9/24/76
3. FREQUENCY OF ISSUE: Quarterly	A. NO. OF ISSUES PUBLISHED ANNUALLY: 4 — B. ANNUAL SUBSCRIPTION PRICE: $5 mem $35 nonmem.
4. LOCATION OF KNOWN OFFICE OF PUBLICATION (Street, City, County, State and ZIP Code) (Not printers): 25 Technology Park, Norcross, GA 30092	
5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS (Not printers): 25 Technology Park, Norcross, GA 30092	
6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR	
PUBLISHER (Name and Address): American Institute of Industrial Engineers, James F. Wolbrink 25 Technology Park, Norcross, GA 30092	
EDITOR (Name and Address): James F. Wolbrink same as above	
MANAGING EDITOR (Name and Address):	

7. OWNER (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)

NAME	ADDRESS
American Institute of Industrial Engineers, Inc.	25 Technology Park Norcross, GA 30092

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES (If there are none, so state)

NAME	ADDRESS
None	None

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES (Section 132.122, PSM)

The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes (Check one)

☒ HAVE NOT CHANGED DURING PRECEDING 12 MONTHS ☐ HAVE CHANGED DURING PRECEDING 12 MONTHS (If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED (Net Press Run)	4600	4700
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	–	–
2. MAIL SUBSCRIPTIONS	3799	3908
C. TOTAL PAID CIRCULATION (Sum of 10B1 and 10B2)	3799	3908
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	89	74
E. TOTAL DISTRIBUTION (Sum of C and D)	3888	3982
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	712	718
2. RETURNS FROM NEWS AGENTS	–	–
G. TOTAL (Sum of E, F1 and 2 - should equal net press run shown in A)	4600	4700

11. I certify that the statements made by me above are correct and complete. SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: James F. Wolbrink

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES (Section 132.121, Postal Service Manual)

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER: James F. Wolbrink

PS Form 3526 Apr. 1974 (Page 1) (See instructions on reverse)

AIIE Research Abstracts

edited by
OWEN MILLER
SENIOR MEMBER, AIIE
University of Missouri

■ These research abstracts represent a partial listing of recently performed research related to industrial engineering. This information was collected by the AIIE Research information Committee from abstracts furnished by various organizations conducting research. It does not represent an exhaustive listing, for not all research has been reported. Additional abstracts for current research will be published on a regular basis. Previous publications of research abstracts appeared in the following issues of *The Journal of Industrial Engineering:* July–August 1958; May–June 1959; January–February 1961; January–February 1962; May–June 1963; January 1966; and February 1967. In addition, research abstracts are published in each issue of *AIIE Transactions*, starting with Volume I, No. 1 (March 1969). The Research Information Committee has made no attempt to judge the merits of any research report. It has only collected and classified the reports.

Copies of original research reports may be obtained by contacting directly the author or organization mentioned in the abstract. Many universities cooperate with University Microfilms, Inc., 300 N. Zeeb Road, Ann Arbor, Michigan 48106, for the publication of masters and doctorate theses. Requests for reproduction of theses could be directed to University Microfilms. Other universities make theses available through Interlibrary Loan. Requests for this service must be made through a library for a short-term loan or reproduction of the thesis.

The Committee wishes to thank the organizations that have helped with this project and to encourage the future participation of all organizations.

Send abstracts to:

Dr. Owen Miller, Chairman
AIIE Research Information Committee
Department of Industrial Engineering
University of Missouri
Columbia, Missouri 65201

The American Institute of Industrial Engineers' classification and codification system consists of twelve primary research areas broken down into a numerical coding system composed of four words. The first word consists of four digits and relates to the research area. The first two digits consist of the general research area into which the abstracts fall and the second two digits the specific area under the general area of interest. The second word consists of three digits and indicates the industry classification to which the abstract is most applicable. Most abstracts will be classified under the classification of 000 since there is universal applicability for the material presented. The third word contains two digits and indicates the source from which the abstract was received. The fourth word contains two digits and indicates the place where the original research document may be obtained. All abstracts have the four-word codification as part of their identification. Some abstracts may contain more than one code as the research activity may fit into several basic research classification areas. Some organizations have punched the code and the title of the research abstracts and are using them as a ready reference for research information. It is quite easy to refer to the codification sheet and select the code numbers of interest. This information can be sorted from the card deck and all research in that area tabulated. ■

XXXX: Research Areas of IE

0100: *Work Measurement*
0105: Procedures Preparatory to Work Measurement
0110: Timing
0115: Measuring Operator Performances
0120: Measuring Job Difficulty
0125: Delay Allowances
0130: Machine Control and Interference Allowances
0135: Personal Allowances and Fatigue
0140: Basic Motion Standard Data
0145: Elemental Standard Data
0150: Control of Standard Times
0155: Special Labor Category Standard Times
0160: Work Sampling
0165: Micro-measurements

0200: *Methods*
0205: Non-repetitive Work Analysis Techniques
0210: Repetitive Work Analysis Techniques
0215: Basic Principles Investigation
0220: Evaluation Procedures
0225: Testing Procedures
0230: Installation Procedures
0235: Operations Design

0300: *Facility Planning*
0305: Plant Layout Analysis Techniques
0310: Plant Layout Basic Principles
0315: Plant Layout Evaluation Procedures
0320: Plant Layout Testing Procedures
0325: Plant Layout Presentation Procedures
0330: Materials Handling Analysis Techniques
0335: Materials Handling Basic Principles
0340: Materials Handling Evaluation and Selection Procedures
0345: Materials Handling Testing Procedures
0350: Materials Handling Equipment Design
0355: Maintenance
0360: Plant Location
0365: Plant Design and Construction
0370: Safety

0400: *Applied Psychology*
0405: Motivation and Psychological Aspects
0410: Physiological Aspects
0415: Sociological Aspects
0420: Biomechanics
0425: Arbitration
0430: Collective Bargaining
0435: Labor
0440: Training
0445: Learning
0450: Job Evaluation
0455: Wage Incentive Plans
0460: Selection of Personnel

0500: *Engineering Economics*
0505: Replacement Theory
0510: Amortization Theory
0515: Interest Rate Concepts
0520: Decision-making

0600: *Organization Planning*
0605: Decentralization Concepts
0610: Social Scientist's Aspects
0615: Control Designs
0620: Communications
0625: Information Theory
0630: Organization Design
0635: Financial Planning

0700: *Materials Processing*
0705: Tool Engineering
0720: Process Research
0730: Mechanization & Automation
0740: Inspection
0745: Product Research

0800: *Production Planning and Control*
0805: Inventory Control
0810: Forecasting
0815: Scheduling
0820: Dispatching
0825: Routing
0830: Packaging
0840: Analysis and Measurement Techniques for Planning and Control
0845: Receiving and Shipping
0850: Stockkeeping and Warehousing
0855: Control Models and Simulation

0900: *Data Processing Systems Design*
0905: Digital Computers
0910: Analog Computers
0915: Accounting Equipment
0920: Desk Equipment
0925: Auxiliary Equipment
0930: Analysis Techniques
0935: Information Retrieval

1000: *Applied Mathematics*
1005: Mathematical Programming
1010: Waiting LIne Theory
1015: Simulation Concepts
1020: Game Theory
1025: Monte Carlo Concepts
1030: Probability Theory
1035: Sampling Concepts
1040: Acceptance Sampling
1045: Quality Control Techniques
1050: Experimental Design
1055: Statistical Theory

1100: *Costs and Cost Control*
1105: Accounting Concepts
1110: Inspection
1115: Budgets
1120: Contol Procedures
1125: Purchasing

1200: *Industrial Engineering Education*
1205: Curricula Undergraduate
1210: Curricula Graduate
1215: Testing
1220: Education Attainment
1225: Business Games
1230: Classroom Schedule Forecasting
1235: Planning

:XXX: Industrial or Business

:000: *General or Nonclassifiable*
:100: *Agricultural, Forestry and Fisheries*
:200: *Mining*
:300: *Contract Construction*
:400: *Manufacturing*
:500: *Transportation, Communication, Electric, Gas and Sanitary Services*
:600: *Wholesale and Retail*
:700: *Finance, Insurance, and Real Estate*
:800: *Services*
:810: *Hospital*
:900: *Government*

:XX: Source Classification

:10: *Unknown*
:20: *College of University Master's Thesis*
:30: *College or University Doctorate Thesis*
:40: *Industrial*
:50: *Government*
:60: *Foundation*
:70: *College or University Bachelor's Thesis*
:80: *College or University Research–No Degree Fulfillment*

:XX: Publication Classification

:10: *Unpublished*
:20: *University Library*
:30: *Technical Journal*
:40: *Trade Journal*
:50: *Book*
:60: *Pamphlet*
:70: *Industrial Library*
:80: *Foundation Library*

1005:000:80:10. *A Mathematical Basis for Integrating Simulation Into Mathematical Optimization;* Hartfiel, D. J.; Curry, G. L.; September 1974; Industrial Engineering Department, Texas A & M University, College Station, Texas 77843.

The work involved in Part I of these papers concerns a method for finding solutions to the programming problem

$$\min_{X} H[p(x), f(x)]$$

where H and f may be expressed by mathematical functions, but $p(x)$ can only be evaluated, for each x, by some device such as a simulator. This method is based on solving

$$\min_{X} H[\alpha_k, f(x)] \text{ for } x(\alpha_k)$$

by standard mathematical programming techniques and then using the simulator to calculate $p[x(\alpha_k)] = \alpha_{k+1}$. Properties are imposed on p and $H[\alpha, f(x)]$ to guarantee that $p[x(o)]$ is a contractive map so that the iteration seeks out an $\bar{\alpha}$ so that $p[x(\bar{\alpha})] = \bar{\alpha}$. It is shown that, under certain conditions on p and $H[\alpha, f(x)]$, $x(\bar{\alpha})$ is a "good" solution to the optimization problem. Error bounds are also given which measure the accuracy of this solution. Part II of these papers then provides a procedure which can be used, in conjunction with the method of Part I, to obtain exact solutions to the optimization problem.

1015:000:80:10. *Business Games: Some Comparative Research;* Biggs, W. D.; Greenlow, P. S.; March 1976; Department of Management Science and Organizational Behavior, The Pennsylvania State University, University Park, Pennsylvania 16802.

Since the development of the first business game millions of dollars have been invested in designing and using games (or simulations), and their use has spread to many other disciplines. Relatively little effort, however, has been expended researching the educational properties of simulations may differ so widely, follow-up research on specific games seems necessary so as to enrich, confirm, or negate previous findings. In this article, such research on a widely used game, FINANSIM, is reported and directly compared with earlier findings.

1040:000:80:30. *Bayesian Single Sampling Plans for Lifetesting with Truncation of the Number of Failures;* Thyregod, P.; Scand. J. Statist. 2, 1975, pages 61-70.

A procedure is given for the determination of acceptance-sampling plans for exponentially distributed life times under known sampling and decision costs. The application of the procedure is shown in an example.

1070:000:30:20. *Optimum Capital Planning of Reliable Systems;* Malik, M. A. K.; June 1975; University of Microfilms, Inc., Ann Arbor, Michigan 48106.

In this thesis, mathematical models are built for maintaining systems which produce services, and the application of these models to good producing systems and demand forecasting is shown. These models incorporate "age or wear," "chance or wear" (combined distribution), and "chance and wear" (compositive distribution) characteristics. Examples of application are given under two assumptions: first that reliability can be improved through replacement only, and second that reliability can also be improved through maintenance. Contemporary replacement policies are scrutinized. In this respect it is proved that the wear-out phenomenon is the basis of system replacement because this increases maintenance costs due to successively decreasing maintenance intervals and in turn necessitates system replacement. The theory is validated through application to an example.

0115:000:80:30. *Marijuana and Human Performance*; Salvendy, G., and McCabe, G., Jr.; Human Factors, Vol. 17 (3), 1975, pp. 229-235.

Four groups of 10 subjects each, representing different levels of marijuana usage, performed two different psychomotor tasks. One group had never smoked marijuana; one group had smoked marijuana previously but had stopped. The other two groups consisted of habitual smokers of marijuana—one smoked a placebo, and the other smoked marijuana just prior to performing the psychomotor tasks. Consistent patterns of inferior performance were found for the marijuana users on both manipulative and coordination skills.

0115:810:80:30. *Objective Evaluation of Quality in Cavity Preparations*; Schiff, A., Salvendy, G., Root, C., Ferguson, G., Cunningham, P.; Journal of Dental Education, Vol. 39, No. 2, 1975.

Significant improvements in reliability of criteria measures were obtained by using a newly developed instrument to evaluate objectively the quality of cavity preparations in comparison to subjective ratings by the instructors of the quality of cavity preparations.

0115:810:80:30. *Improving Evaluation of Amalgam Restorations;* Salvendy, G., Joost, M., Cunningham, P., Ferguson, G., Wilko, R., and Dees, R.; Journal of Dental Education, Vol. 40, No. 6, 1976.

Dental educators have long recognized the lack of specificity of criteria used in evaluating dental restorations, and much effort has gone into searches for objective criteria. The purpose of this paper is to report on how the reliability of overall quality ratings may be improved by using existing criteria and on how identification of those aspects of quality which contribute the most to the overall ratings may help students and instructors increase training effectiveness.

0300:810:30:10. *The Use of Categorization in the Regional Allocation of Emergency Medical Services;* Salguero, O.R.; January, 1977; Belvier Engineering Library, University of Pittsburgh, Pittsburgh, PA 15261.

Develops a categorization of emergency facilities methodology that is suitable for reducing the duplication of emergency services. A balanced allocation of emergency medical services within a region is obtained by using mismatches between patient's needs and emergency facilities capabilities as a measure of effectiveness. The model considers the dynamics of patient's patterns in selecting an emergency facility.

0355:400:20:10. *A Computerized Maintenance Management Control System;* Mann, L., Bradberry, J., and Potier,

P.; August, 1976; Industrial Engineering Department, Louisiana State University, Baton Rouge, LA 70803.

This research resulted in a comprehensive management information system for maintenance for process plant operations.

0355:500:80:80. *Increasing Efficiency in Bus Maintenance Operations;* Martin-Vega, L. A.; July, 1976; Progress Report TRI-76-01, Transportation Research Institute School of Engineering, University of Puerto Rico, Mayaguez, Puerto Rico, 00708.

The Commonwealth of Puerto Rico Metropolitan Bus Authority is currently carrying out an overall improvement program to upgrade both the level and quality of its service. Management acknowledges that despite the work being carried out in the planning and scheduling aspects of the improvement program the number of buses in service are not adequate. This report presents the progress made in analyzing data collected by the maintenance staff as well as in the development of systems for routine utilization of data in the future.

0405:400:80:30. *Effects of Equitable and Inequitable Financial Compensation on Operator's Productivity, Satisfaction and Motivation;* Salvendy, G.; International Journal of Production Research, Vol., 14, No. 2, pp. 305-310, 1976.

In studying the activities of 274 shop floor hourly operators during a period of 16 months, the evidence indicates that during a period of equitable payment system the number of grievances decreased 31 times, labor turnover 16 times, lateness 6 times and absenteeism nearly 3 times while at the same time productivity increased by 22% in comparison to an equivalent period of inequitable financial incentive payment system. These improvements are attributable to the relative and not absolute financial compensation of the operators.

0410:810:20:20. *Effects of Maternal Variables on Infant Health For the White Mountain Apaches;* Chin, J. C., and Dobson, W. F., Jr.; December, 1974; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

Infant medical, environmental, socio-economic and maternal health data were gathered for a linear discriminant analysis of factors related to infant mortality and morbidity on the reservation.

0440:810:80:30. *Training in Force Exertions during Condensation of Dental Amalgam;* Salvendy, G., Joost, M., Ferguson, G., Cunningham, P., Schiff, A., and Wilko, R.; Journal of Dental Education, Vol. 40, No. 6, 1976.

It has been suggested that a pressure of at least 2,000 pounds per square inch should be applied during certain amalgam condensations to achieve optimal physical properties of the material. If this value is accepted, then a force of at least 2.6 pounds should be applied to the small amalgam condenser (SUNY B – 10, 1.3×10^{-3} square inch area of condenser face) during amalgam condensation. The purpose of this paper is to illustrate the impact of a force analysis and recording instrument on the force used by dental students during amalgam condensation.

0440:810:80:30. *A Second Generation Training Simulator for Acquisition of Psychomotor Skills in Cavity Preparation;* Salvendy, G., Root, C., Schiff, A., Cunningham, P., and Ferguson, G.; Journal of Dental Education, Vol. 39, No. 7, 1975.

Thirty freshmen dental students, without prior dental technique laboratory experience, were divided randomly into two groups of 15 subjects. One group (control) was trained in the traditional way while the other group used a newly developed dental training simulator. The results for Class I cavity preparations in the mandibular right first molar indicate that a two-thirds reduction in training time occurred for the group trained by the simulator as compared to the control group. The mean training times were 57 minutes for the control group and 21 minutes for the experimental group. No statistically significant differences in quality were noted.

0440:810:80:50. *Training Dental Students in Cavity Preparations;* Salvendy, G.; Human Factors in Health Care, edited by R. M. Pickett and T. J. Triggs, D. H. Heath & Co., Lexington, Mass., 1975.

The United States has a long-standing problem: there are too few dentists to care properly for the oral health needs of all her people. There is thus, a joint need to produce more dentists and to make them more productive. A series of experimental evidences are presented here that show how the productivity of psychomotor operations is significantly increased in dentistry by adopting analytical training techniques based on skills analysis. The benefits include a significant reduction of training time for dental students and often an increase in performance quality.

0460:400:80:30. *The Prediction of Operator Performance on the Basis of Performance Tests and Biological Measures;* Salvendy, G., and Stewart, G. K.; AIIE Transactions, December, 1975, pp. 379-387.

Thirty-five female bench assembly operators were utilized as subjects to support the hypothesis that production performance can be better predicted from biological measures plus manipulative test performance scores than just manipulative test performance. The following measures were utilized: 1) for biological data; EMG, heart beat and sinus arrhythmia 2) for test performance; the One-Hole test 3) for criteria measure; production performance. For the hybrid model significant concurrent validities, using shrinkage statistics, of 0.61, 0.76, and 0.69 multiple correlations were obtained with 8, 20, and 30 variables respectively. Sinus arrhythmia at rest was the single most powerful predictor of production work performance with a –0.45 correlation coefficient.

0460:400:80:30. *Selection of Industrial Operators: The One-Hole Test;* Salvendy, G.; International Journal of Production Research, Vol. 13, No. 3, pp. 303-321.

The one-hole test was developed to provide a situation in which unlike that in the Purdue pegboard and similar tests, the reach, grasp, move and position elements were consistent and their rate of improvement measurable. Results support the hypothesis that measurements of rate of learning or of performance after 600 cycles of practice have a higher concurrent and predictive validity of production output than measurements of initial performance on the test. The one-hole test has a mean predictive validity of 0.75 (N = 228) between test scores and rate of learning in the training school; and a mean concurrent validity of

0.66 (N = 307) between test scores and mean production performance of experienced operators. Norms (N = 1422), reliabilities (N = 158) and correlation of the one-hole test scores with anthropometric measurements, personality, intelligence, age and other manual dexterity tests are discussed.

0500:000:20:20. *Cement Production and Consumption in Saudi Arabia (A Study on Prediction and Investment);* Alomair, O. A., and Young, H. H.; May, 1974; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

Several economic factors were investigated through regression and exponential smoothing analysis with only government expenditures and the country's GNP being found to give considerable influence in cement consumption. A decision model for investment under conditions of uncertainty was constructed based on predicted values generated by the exponential smoothing approach. Among the alternatives presented, only 300 and 400 and 500 thousands tons of annual output levels are suggested by the model.

0500:000:80:50. *The Conservation Response;* Dumas, L. J.; The Conservation Response, Lexington Books, D.C. Heath, Lexington, Mass (1976).

Strategies for the energy conserving design and operation of energy-using systems are described and evaluated. Energy use in buildings, transportation, industry and the food system are all considered. The strategies discussed illustrate the enormous potential for reducing energy consumption in the U.S. while at the same time maintaining the standard of living. It is further argued that this approach to the energy problem is perfectly compatible with the goal of improving environmental quality.

0520:000:20:20. *Decision Economics in Multi-Attribute Single and Double Acceptance Sampling for Small Lots;* Djivanovic, A., and Bedworth, D. B.; January, 1975; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

General statistical-economic multi-attribute single and double acceptance sampling models are presented. A computer program is designed to perform complex and long computations incorporated in the models. In addition to optimal costs, the program supplies optimal sampling plan combinations and the expected costs of inspection, rejection, and acceptance.

0520:000:20:20. *Decision Factors in the Development of an Iron and Steel Productive Capability in Saudi Arabia;* Saif, A. A., and Young, H. H.; June, 1975; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

The basic objective of this research was to clarify and suggest solutions for several obstacles facing the establishment of an iron and steel industry in Saudi Arabia. The study provides the essential techno-economic evaluations of factors that could be attributed to the productive capability of the proposed steel plant, decisions and recommendations.

0520:000:30:20. *A Study of Analytic Methods for the Analysis and Design of Decision Policies in Ind. Dynamics Models;* Doerer, H. T., and Bedworth, D. D.; July, 1974; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

The objective of this dissertation is to develop and demonstrate analytic methods for the analysis and design of decision policies in industrial dynamics models. Three areas of dynamic systems theory are investigated: stability theory, sensitivity theory, and optimal control theory. Concerning the problem of determining the feedback decision policies that optimize a given performance functional, three types of solutions are considered: (1) dynamic programming, (2) optimal regulator theory, and (3) perturbation feedback theory.

0520:000:30:20. *A Decision Theory Approach to Personnel Management;* Tillman, J. B., and Smith, R. L.; May, 1975; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

This dissertation investigates the application of decision theory models to personnel management decision situations. It includes a taxonomy of personnel management decision situations. It identifies and describes decision theory models which might be applied to personnel management decision situations. A systematic method of analysis which can be used as a guide by the personnel manager in the application of decision theory methods to practical personnel management problems is developed.

0520:900:20:20. *Decision Analysis and Competitive Bidding in Government Contracting;* Flynn, D., and Hoyt, C. D., Jr.; December, 1974; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

Aspects of decision analysis and bidding strategy models were applied to construction contracting from the viewpoint of the contracting agency. The Gates' analysis of bid spreads was found to provide a valuable decision tool to protect the owner's interests in competitive bidding situations.

0630:400:80:10. *Coordinating Mechanisms between R&D and Marketing;* Souder, W. E.; August, 1977; Industrial Engineering Department, University of Pittsburgh, Pittsburgh, PA 15261.

The purposes of this study are to define the types of collaboration problems that exist between R&D and marketing at 18 different companies, and to examine various ways for handling these problems. Task force, matrix and other organizational methods are being studied.

0700:900:20:20. *Solid Waste Heat Reclamation Feasibility Study Williams Air Force Base, Arizona;* Laska, L. M., and Wilson, L. F.; June, 1975; Department of Industrial Engineering, College of Engineering and Applied Sciences, Arizona State University, Tempe, AZ 85281.

This study evaluates the solid waste heat reclamation processes available now for use on military bases and compares them to present solid waste generating conditions at Williams Air Force Base, Arizona. All the processes studied were found to be more expensive to operate, on a cost-per-ton-of-solid-waste basis, than is the current practice of sanitary landfilling.

Letter to Editor

Editor:

An error was made in a previous article of mine in *AIIE Transactions.* The article is "Transportation Routing Problem – A Survey" that appeared in Volume 6, No. 4, Pages 288-301.

In the article, I presented an integer programming formulation of the vehicle routing problem. Dr. Bruce Golden of the University of Maryland – College of Business and Management recently developed a counter example such that solution of my formulation would yield subtours.

The counter example is "Consider the situation where $n = 3, m = 1, C_1 = 100, D_2 = 30, D_3 = 40, P_1 = 1, d_{12} = d_{21} = d_{13} = d_{31} = 50, d_{23} = d_{32} = 10$. The solution to the formulation is $y_{23} = y_{32} = 1$ and $x_{23} = .4, x_{32} = .3$ which corresponds to a subtour."

A correct formulation will appear in "Implementing Vehicle Routing Algorithms" by Golden, Magranti, and Nguyen forthcoming in *NETWORKS.*

I offer my apologies to *Transactions* and its readers but please note that integer programming formulations are seldom if ever used to solve vehicle routing problems. Also, I would like to congratulate Dr. Golden on the professional manner in which he handled this correction.

Wayne C. Turner
Associate Professor
Oklahoma State University
Stillwater, OK 74074

Editor:

In responding to questions from several readers concerning my recent article "Allocation of Interstage Storage Along an Automatic Production Line" in *AIIE Transactions,* March 1976, Vol. 8, No. 1, pp. 146-152, I have noticed that equation (3) on page 148 of that article was printed incorrectly. The correct form of equation (3) for the steady state production rate of a 2-stage line is

$$P(\text{line=U}) = (1-P2) - (1-P2)/\left[1+(1/P2)\sum_{i=1}^{N_1}\{(1-P1)P2/[P1(1-P2)]\}^i\right]. \qquad (3)$$

I apologize to readers for this error and thank them for their interest in my results.

Theodore J. Sheskin, Assistant Professor
Industrial Engineering Department
Fenn College of Engineering
The Cleveland State University
Cleveland, Ohio 44115

CALL FOR PAPERS

Kansas City, Missouri • 1977 SYSTEMS ENGINEERING CONFERENCE • November 2-4, 1977

You are invited to submit an abstract of a paper for presentation at the American Institute of Industrial Engineers' 1977 Systems Engineering Conference to be held in Kansas City, Missouri, November 2-4, 1977, at the Radisson Muehlebach Hotel. The Conference will be hosted by the Kansas City Chapter.

Each Annual Systems Engineering Conference is theme-oriented, so papers should contribute to the over-all objectives of the Conference theme, "*Productivity: The Heart of America.*" The papers should deal with current state-of-the-art techniques and applications. Suggested topical areas would include:

Materials Requirement Planning (MRP)
Manufacturing Technology
Productivity
IE Information for Management Control
Applications of Mini-Micro Computers

In addition to the Theme Sessions, DELPHI Sessions will be conducted to provide speakers with an alternative forum for communicating their ideas to other attendees. You are also invited to submit an abstract for one of these informal sessions.

In either case, three (3) copies of a 300-400 word abstract must be submitted to arrive no later than April 15, 1977. Send to:

1977 Systems Engineering Program
c/o Conference Department
American Institute of Industrial Engineers
25 Technology Park/Atlanta
Norcross, Georgia 30092

Authors of those papers selected for presentation will receive notification of acceptance by June 1, 1977. This notification will include detailed instructions on preparation of the final manuscript that will be required by August 15, 1977.

International Journal of

Production Research

Contents for Volume 14, No. 6 November 1976

A stochastic explanation of some structural properties of a ground surface.
R. S. Sayles and T. R. Thomas

Measurement of tool-work interface temperature in hot machining.
N. N. S. Chen and C. F. Ho

Coordinating control of production inventory systems.
S. Axsäter

Production scheduling of perishable products.
J. P. Monahan

A viable alternative to the 'method of stages' solution of series production systems with Erlang service times.
N. P. Rao

The location of variable work stations and the performance of production flow lines.
C. A. Carnall and R. Wild

Failure characteristics of multi-spindle multi-station complex machines.
C. Jain and W. M. Hancock

Multiple deferred state sampling inspection.
A. W. Wortham and R. C. Baker

Book reviews.

Published Bi-monthly. Subscription rates (payable in advance): USA, Canada and Mexico (air-freight) $70.00. Single copy $8.80 postage extra. Privilege rate to Members of The American Institute of Industrial Engineers (air-freight) $25.00. Subscribers paying in U.S. dollars should pay their cheques into our account in the USA. Cheques must be drawn in favour of Taylor & Francis Ltd, Account no. 41-682-132, and sent with the order to Taylor & Francis Ltd, P.O. Box 9137, Church Street Station, New York, N.Y. 10049 (for our account with the Bankers Trust Company).

Joint Sponsors

The Institution of Production Engineers

The American Institute of Industrial Engineers Inc.

Published by

TAYLOR & FRANCIS LTD.

10-14 Macklin Street • London WC2B 5NF

AIIE is a cosponsor of *The International Journal of Production Research.* Manuscripts submitted for possible publication should be sent to: The Editor, *International Journal of Production Research,* Department of Engineering Production, PO Box 363, The University of Birmingham, Birmingham B15 2TT England.

1977's single most important PRODUCTIVITY event is really two great events...

The 28th ANNUAL CONFERENCE & CONVENTION

May 24~27, 1977

"PRODUCTIVITY, PROGRESS, PROFITS" is this year's theme. It would be difficult to find three elements more vital to the effective operation of any organization in business, industry, or government.

The expertise of key figures from leading corporations, government service, and educational institutions will cover the gamut of IE specialization. And the program will emphasize both theory and practice. Practical applications of problem-solving techniques will demonstrate how concepts can be translated into concrete on-the-job solutions.

A special executive panel of experts from the military and private sectors will discuss the pros and cons of the "U.S. Air Force Military Standard on Work Measurement." Panelists are: Brigadier General Hans H. Driessnack, U.S. Air Force Systems Command; William Waddell of General Electric; Gordon E. Sylvester of General Dynamics, and Donald E. Moore, Kirtland Air Force Base.

Three outstanding corporate presidents will analyze the impact of industrial engineering in their organizations:

J. M. HAGGAR, JR., president of Haggar Co., Dallas, discusses "The Role of the Industrial Engineer in a Highly Competitive Sales-Oriented Industry."

GENE LEVINE, president of Gene Levine Associates of Dallas, speaks on "Cashing in on the Golden Opportunities of Engineering."

JACK C. VANDER WOUDE, president and general manager of Texas Eastman Co., Longview, looks at "Management for Greater Productivity."

★ Sponsored by the American Institute of Industrial Engineers

The annual PRODUCTIVITY ENGINEERING SHOW

May 24-26, 1977

This show provides the "other half" of the PRODUCTIVITY equation: the tools, materials, and systems that turn concepts into reality. The products and services displayed at the Productivity Engineering Show provide immediate access to the latest technology available for turning the ideas learned at the Conference into solutions that will work for your organization.

As many as 70 exhibitors will be bringing new products and systems to the attention of Conference attendees.

Some of the PRODUCTIVITY areas being represented include:

- Material Handling Equipment
- Production Machinery
- Controls & Instrumentation
- Work Measurement Devices
- Computers & Software
- Plant/Office Furniture & Equipment
- Packaging Equipment
- Consulting & Professional Services

...designed for every engineer, corporate executive, or manager involved in any aspect of PRODUCTIVITY: the dominant factor in the economic well-being of U.S. business and industry-today and tomorrow. May 24-27 at the Fairmont Hotel in Dallas, Texas

This year's diverse technical program is ideal for every practicing Industrial Engineer. Look at this sampling of topics:

"Before Buying MRP: Stop—Look—Listen!"
"Productivity—People Participation"
"Indirect Labor Control in a Mail Order Business"
"Managing the Courts for Productivity"
"The Law Says: End Race Discrimination Now! And It Means It!"
"A Pre-Surgical Screening Program: An Effort to Reduce Soaring Health Care Costs"
"An Economic Analysis of a Plastic Recycling Program"
"Maintenance Incentive Plans"
"Small is Beautiful—the Volvo Experiment"
"Top Team Planning for Higher Productivity"
"Low-Cost Measurement of Indirect Labor Productivity"

Plant Tours provide an opportunity to see industrial engineering in action—at the Dallas Bulk Mail Center and at the American Airlines Flight Academy.

The annual Chief IE Symposium offers the ideal setting for an invaluable interchange of ideas among managing IE's. This year's topics include:

PLUS luncheon speakers, receptions, and a special Spouse's Program—all designed to create an environment that encourages a sharing of viewpoints among professionals with similar interests.

If you're concerned with PRODUCTIVITY, this conference should be "Top Priority." Send today for complete information. Or call (404) 449-0460.

Registrar, The 28th Annual AIIE Conference
25 Technology Park/Atlanta, Norcross, GA 30092

Dear AIIE:
I'm vitally interested in PRODUCTIVITY. Please rush full data on your program and registration for the 28th Annual Conference and Convention in Dallas, Texas.

Name _____
Company _____
Street _____
City _____ State _____ Zip _____
Telephone _____

A SPECIAL OPPORTUNITY FOR EXHIBITORS, this Show provides a pre-interested market—thousands of engineers and managers looking for better ways to solve the complex PRODUCTIVITY problems facing their organizations.

If your product or service can increase PRODUCTIVITY or improve profitability, this Show offers an unparalleled opportunity to meet your market face-to-face.

THE BOTTOM LINE FOR EXHIBITORS is a well-organized exhibit program and exceptionally high-quality traffic. The IE's who buy will be in Dallas. So should you!

If you market products or services to productivity-minded engineers and managers, you cannot afford to miss this Show. Exhibit space is still available—if you act now! Write or call (404) 449-0460.

Manager, The Productivity Engineering Show
25 Technology Park/Atlanta, Norcross, GA 30092

Dear AIIE:
We want to reach engineers and managers at the annual Productivity Engineering Show in Dallas, Texas. Please send full details immediately on exhibiting at the show.

Name _____
Company _____
Street _____
City _____ State _____ Zip _____
Telephone _____

Development and Applications Paper Award

In the interest of encouraging the publication of outstanding development and application work, the American Institute of Industrial Engineers, Inc. is pleased to announce its sponsorship of the second annual *AIIE Transactions* Development and Applications Paper Award. The amount of the award will be $2500.

1) ***Procedure for Submitting Papers:*** Papers submitted to *AIIE Transactions* during the time period of January 1, 1977, through December 31, 1977, shall be considered for the 1977 Award. Authors wishing to designate specifically that their papers be considered for the Award, as well as for publication in *AIIE Transactions,* are advised to do so by having the working Development and Applications Paper typed at the top of the title page of manuscripts submitted. Further details on submitting papers are given on the inside back cover page of each issue of *AIIE Transactions,* and should be referred to. Papers must be accepted for publication in *AIIE Transactions* to become eligible for the Award.

2) ***Eligibility for Competition:*** Anyone, with the exception of AIIE employees, AIIE Board of Trustees members, and members of the Award Selection Committee, may submit a paper for consideration–authors need not be members of AIIE.

3) ***Procedure for Judging Papers:*** All papers considered for the Award shall be evaluated by an Award Selection Committee whose members shall be appointed for specific terms by the Editor-in-Chief. The Selection Committee shall present its recommendations to the Vice President-Publications for final approval. Should the Selection Committee find no paper worthy of a prize, no prize will be awarded that year.

4) ***Criteria for Judging Papers:*** Criteria for judging papers shall include, but not be limited to, the following: professional relevance, importance of the problem considered, impact of the work upon the profession, savings and benefits resulting from the work, and endorsement of the work by its users.

Authors should be prepared to provide, as requested, the following:

a) A statement of the monetary amount of savings, or other benefits, resulting from their work, together with an independent certification of the authenticity of the statement from an appropriate organization or unit, such as an accounting department.

b) Names and addresses of users, in positions of responsibility, willing to provide endorsement of the utility of the author's work.

It is expected that papers considered will be well written, and well documented. A paper should, ideally, include a discussion of (a) the definition and context of the actual problem, (b) the current approach to solving the problem and its deficiencies, (c) the development of the model (and why others were rejected), (d) the data collection and/or generation, with a discussion of the questions of experimental design (when appropriate), data reliability, and the assumptions made, (e) model validation, (f) the solution arrived at–its form and content, (g) sensitivity and analysis of the model to test assumptions, (h) model implementation, and robustness of solutions in real life application, (i) conclusions and recommendations. These are ideal desiderata, and constitute a goal rather than a checklist. It is recognized that in some application areas not all the foregoing desiderata will apply: papers dealing with such applications will be considered on their own merits.

5) ***Confidential Information:*** If any information in a submitted paper has been altered or otherwise coded to maintain confidentiality, a statement to this effect is required in the letter of submittal; the omission of such a statement will be considered to be a representation by the author, or authors, that no information has been altered or coded. If coding of data is necessary, it is preferable that the coding be done in such a way as to provide an accurate picture of relative results.

6) ***Deadline for Submission:*** December 31, 1977. The Award winner will be announced in the summer of 1978.

AIIE TRANSACTIONS

■ *AIIE Transactions*, subtitled "Industrial Engineering Research and Development" is devoted primarily to the publication of new developments and research in industrial engineering—to foster a rich exchange of ideas among researchers in the field and promulgate to the industrial engineering community new research and development ideas. *AIIE Transactions* does not desire to compete with other professional journals for manuscripts which are clearly in their special fields of interest. Papers on applications are strongly encouraged. Papers are expected to be of demonstrable relevance to industrial engineering.

Authors are encouraged to include, where applicable, one or more meaningful examples as an integral part of their manuscripts.

Good state-of-the-art survey papers are encouraged. Short manuscripts with significant content are refereed for possible publication as Technical Notes.

AIIE Transactions include a "Reader Comments" section. Correspondence from readers which is pertinent to the contents of *AIIE Transactions* is encouraged.

Manuscripts are refereed and the identity of the author(s) will not be revealed to referees during the review process. The identities of the referees will not be revealed to the author(s). However, reviews of each manuscript will be exchanged among referees.

The December issue each year contains a sample Reviewer's Check List. Authors should consult this for a guide to the specific points reviewers analyze in making their evaluations.

Four copies of the manuscript should be submitted to the Editor-In-Chief, *AIIE Transactions*, American Institute of Industrial Engineers, 25 Technology Park, Norcross, Georgia 30092. The author's letter of transmittal should state that the material has not been copyrighted, published, or submitted for publication elsewhere. Please also suggest the appropriate Department Editor to handle reviews of the paper (see instructions in front).

Articles published in recent issues of *AIIE Transactions* will serve as examples of the types of manuscripts and styles which are appropriate. Particularly important requirements are listed below.

General—Keep all contributions as brief as possible consistent with clear, concise writing. In all contributions, summarize or cite, do not repeat, arguments already available elsewhere, and make references to related previous work. The Editor-In-Chief will give priority to short papers over long ones. For other than exceptional papers, authors submitting over 20 typewritten pages can anticipate publication delays.

All manuscripts submitted for consideration should be typed on 8½ x 11 inch paper and should be double-spaced throughout, including references, footnotes and abstracts. Great care must be taken to assure that the manuscript is typographically correct.

Title Page—A separate page should contain the title of the manuscript, together with the name, AIIE membership grade, company or institutional affiliation, and complete mailing address of each author, given in the order in which the listings should appear in the published version.

Abstract—Each article and Technical Note must be accompanied by a nonmathematical abstract of about 100 words, suitable for literature retrieval purposes. The abstract should summarize the purpose, scope and principal results and conclusions. Literature citations should not be made in the abstract.

Footnotes—The use of footnotes must be kept to a minimum because they make editorial layout more difficult. When employed, footnotes should be indicated by consecutive superscript numbers in the text. The footnotes should be listed on a separate page of the manuscript.

Mathematical Expressions—Use the solidus whenever possible in preference to built-up fractions, write complicated exponentials in the form exp ( ), avoid subscripts or superscripts on subscripts or superscripts, and, in general, minimize unusual or expensive typographical requirements. Make subscripts and superscripts large and clear; label Greek letters and unusual or ambiguous symbols on first occurrence, such as subscript "zero" to distinguish it from the letter "O." Be sure that it is clear if letters in formulas are capital or lower case. Equations should be numbered consecutively with the numbers in parentheses on the same line as the equation, flush right.

Figures—Illustrations are called figures and are numbered consecutively with Arabic numerals. Each should appear on a separate page of the manuscript and be cited in the text. Figure captions should be listed on a separate page.

Figures must be professionally prepared. One set must be of high quality, suitable for reproduction. Use black ink on tracing paper and prepare the artwork in a size larger than it will be when printed. Sketches are not acceptable. All artwork must be complete and correct.

When appropriate, photographs are acceptable if they are sharp, glossy prints with good contrast.

References—References should be listed alphabetically by last name of author on a separate sheet at the end of the manuscript. References do not appear as footnotes. They should be numbered in brackets and the proper numbers should be used in brackets when references are cited in the text. Example:

[1] Saaty, Thomas L., *Mathematical Methods of Operations Research*, McGraw-Hill Book Company, New York, 132-139 (1959).

[2] White, John A. and Francis, Richard L., "Solving a Segregated Storage Problem Using Branch-and-Bound and Extreme Point Ranking Methods," *AIIE Transactions*, **3** 1, 37-44 (March 1971).

References should be to readily-available books and publications; references to reports of limited circulation must be avoided. All references must give complete, clear information.

Biographical Sketch—Brief biographical sketches of each author or co-author should accompany the manuscript. Each sketch should include: the author's full name, title and organization; research activities (interests) and/or responsibilities in which the author is currently engaged; relevant former positions; educational background; AIIE membership grade; other professional societies with which the author is affiliated; and any additional information considered pertinent by the author.

Alterations—Page proofs will be sent to the author for checking. The author should make only those corrections required to make the article grammatically and technically correct. A limited number of alterations after type has been set are unavoidable, but the cost of making extensive alterations in proof as a result of an inaccurate or unclear manuscript will be charged to the author.

Voluntary Page Charges—After a manuscript has been accepted for publication, and at the time page proofs are transmitted to the author for checking, the author's company or institution will be requested to pay a voluntary charge of $40 per printed page to cover part of the cost of publication. Page charges for *AIIE Transactions*, like those of other professional societies, are not obligatory, nor is their payment a prerequisite for publication. The author (or senior author) will receive 100 free reprints if the charge is honored. Detailed instructions will accompany the galley proof.

601283 E
XEROX UNI_
300 NORTH ZEEB RD
ANN ARBOR MICH 48106